

**ВІННИЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ**

Факультет комп'ютерних систем та автоматики  
Кафедра комп'ютерних систем управління  
Спеціальність 151 Автоматизація та комп'ютерно-інтегровані технології  
Освітньо-професійна програма Інтелектуальні комп'ютерні системи

**ЗАТВЕРДЖУЮ**

Завідувач кафедри КСУ  
Дубовой В.М.

« \_\_\_ » \_\_\_\_\_ 2019 року

**МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА**

Розробка методів оцінки ефективності автоматичного перекладу мови в  
інтелектуальних системах автоматизації  
08-01.МКР.003.00.000

Студент групи 2АКІТ-18м Власюк М.В.

Керівник к.т.н., доцент Биков М.М.

Рецензент к.т.н., доцент Севастьянов В. М.

# ВІННИЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

Факультет комп'ютерних систем та автоматики

Кафедра комп'ютерних систем управління

Освітньо-кваліфікаційний рівень магістр

Спеціальність 151 Автоматизація та комп'ютерно-інтегровані  
Освітньо-професійна програма Інтелектуальні комп'ютерні системи

## ЗАТВЕРДЖУЮ

Завідувач кафедри КСУ

Дубовой В.М.

« 2 » 09 2019 року  
Протокол № 1 засідання кафедри КСУ  
від 2.09.2019р.

## З А В Д А Н Н Я НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Власюку Максиму Володимировичу

(прізвище, ім'я, по батькові)

1. Тема магістерської кваліфікаційної роботи «Розробка методів оцінки ефективності автоматичного перекладу мови в інтелектуальних системах автоматизації»

керівник магістерської кваліфікаційної роботи Биков Микола Максимович, к. т. н., доцент

( прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затвержені наказом вищого навчального закладу від «02» 10 2019 року № 254

2. Строк подання студентом магістерської кваліфікаційної роботи 10.12. 2019 року

3. Вихідні дані до магістерської кваліфікаційної роботи текст на англійській мові без формул, абревіатур і скорочень обсягом не менше 60 слів, тексти перекладені людиною та програмами машинного перекладу Pragma, PROMT, Google Translate, середовища моделювання – Visual Studio 2017.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити) Вступ (актуальність проблеми, постановка задачі); Варіантний аналіз існуючих методів і підходів до вирішення задачі оцінювання систем машинного перекладу мови; Розробка вдосконалених методів оцінювання систем машинного перекладу мови; Розробка алгоритмів та програмного забезпечення; Економічна частина.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень) 1. Математична модель процесу перекладу тексту – 2 плакати ; 2.Схема алгоритму приведення тексту до зручного для оцінювання вигляду – 1 аркуш ; 3. Схема алгоритму порівняння двох текстів – 1 аркуш; 4. Структура програми-аналізатора; 5. Файлова структура головної програми –1 аркуш; 6. Результати машинного експерименту – 10 плакатів.

## 6. Консультанти розділів магістерської кваліфікаційної роботи

| Розділ змістової частини роботи | Прізвище, ініціали та посада консультанта | Підпис, дата   |                  |
|---------------------------------|-------------------------------------------|----------------|------------------|
|                                 |                                           | завдання видав | завдання прийняв |
| 1                               | к.т.н, професор кафедри КСУ Биков М.М.    |                |                  |
| 2                               | к.т.н, професор кафедри КСУ Биков М.М.    |                |                  |
| 3                               | к.т.н, професор кафедри КСУ Биков М.М.    |                |                  |
| 4                               | к.е.н, доцент кафедри ЕПВМ Ратушняк О. Г. |                |                  |

## Календарний план

| № з/п | Назва етапів Роботи                                                                                                                                | Строк виконання етапів роботи | Примітка |
|-------|----------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|----------|
| 1     | Вступ, варіантний аналіз існуючих методів і підходів до вирішення задачі оцінювання систем машинного перекладу мови, обґрунтування обраного методу | 02.10.2019р.                  |          |
| 2     | Розробка вдосконалених методів оцінювання систем машинного перекладу мови; Розробка алгоритмів та програмного забезпечення;                        | 20.10.2019 р.                 |          |
| 3     | Розробка алгоритмів та програмного забезпечення                                                                                                    | 02.11.2019 р.                 |          |
| 4     | Проведення експериментальних досліджень.                                                                                                           | 15.11.2019 р.                 |          |
| 5     | Підготовка економічної частини                                                                                                                     | 22.11.2019 р.                 |          |
| 6     | Оформлення пояснювальної записки, графічного матеріалу і презентації                                                                               | 30.11.2019 р.                 |          |
| 7     | Апробація результатів дослідження. Публікації                                                                                                      | 06.12.2019 р.                 |          |
| 8     | Попередній захист                                                                                                                                  | 10.12.2019 р.                 |          |
| 8     | Захист МКР                                                                                                                                         | 12.12.2019 р.                 |          |

Дата видачі завдання “ 02 ” 09 2019 року

Студент \_\_\_\_\_ Власюк М. В.

Керівник магістерської кваліфікаційної роботи \_\_\_\_\_ Биков М.М.

## ЗМІСТ

|                                                                                                                                |    |
|--------------------------------------------------------------------------------------------------------------------------------|----|
| АННОТАЦІЯ .....                                                                                                                | 6  |
| ВСТУП.....                                                                                                                     | 7  |
| 1 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ І ПІДХОДІВ ДО ВИРІШЕННЯ ЗАДАЧІ<br>ОЦІНЮВАННЯ СИСТЕМ МАШИННОГО ПЕРЕКЛАДУ МОВИ.....                    | 11 |
| 1.1 Аналіз об'єкта дослідження .....                                                                                           | 11 |
| 1.2 Класифікація систем МП .....                                                                                               | 13 |
| 1.3 Проблеми побудови систем МП .....                                                                                          | 15 |
| 1.4. Аналіз існуючих принципів функціонування та методів оцінювання систем МП<br>.....                                         | 20 |
| 1.5 Аналіз недоліків існуючих методів оцінювання якості машинного перекладу ....                                               | 36 |
| 1.6 Висновки.....                                                                                                              | 37 |
| 2 РОЗРОБКА ВДОСКОНАЛЕНИХ МЕТОДІВ ОЦІНЮВАННЯ СИСТЕМ<br>МАШИННОГО ПЕРЕКЛАДУ МОВИ.....                                            | 39 |
| 2.1. Модифікований метод N-грам із зміненою точністю та використанням морфем.                                                  | 39 |
| 2.2. Метод оцінки ефективності систем МП на основі інформаційно-статистичного<br>критерію і моделі потенціальної системи ..... | 45 |
| 2.3. Висновки.....                                                                                                             | 50 |
| 3 РОЗРОБКА АЛГОРИТМІВ І ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ.....                                                                          | 51 |
| 3.1 Розробка алгоритмів для реалізації модифікованого методу порівняння кількості<br>збігів N-грам.....                        | 51 |
| 3.2 Дослідження систем машинного перекладу розробленим методом .....                                                           | 55 |
| 3.3 Дослідження адекватності розроблених математичних методів і моделей .....                                                  | 59 |
| 3.4 Інструкція користувачеві .....                                                                                             | 62 |
| 3.5 Висновки.....                                                                                                              | 62 |
| 4 ЕКОНОМІЧНА ЧАСТИНА.....                                                                                                      | 64 |
| 4.1 Оцінювання комерційного потенціалу розробки .....                                                                          | 64 |
| 4.2 Прогнозування витрат на виконання науково-дослідної роботи .....                                                           | 67 |
| 4.3 Оцінка внеску НДР .....                                                                                                    | 71 |
| 4.4 Висновок.....                                                                                                              | 73 |
| ВИСНОВКИ .....                                                                                                                 | 75 |

|                                                     |    |
|-----------------------------------------------------|----|
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ .....                    | 77 |
| ДОДАТКИ .....                                       | 84 |
| Додаток А. ТЕХНІЧНЕ ЗАВДАННЯ .....                  | 85 |
| Додаток Б. Лістинги основних модулів програми ..... | 89 |
| Додаток В. ПЕРЕЛІК_ГРАФІЧНИХ МАТЕРІАЛІВ.....        | 97 |

## АННОТАЦІЯ

Магістерська дисертація присвячена розв'язанню актуальної проблеми підвищення ефективності методів оцінювання систем автоматичного перекладу іншомовних текстів.

В результаті проведених досліджень було запропоновано нову модель процесу перекладу мови, вдосконалено метод статистичного оцінювання систем автоматичного перекладу мови, розроблено методику автоматизованого оцінювання систем машинного перекладу мови на основі запропонованих методів. Розроблено програмне забезпечення для реалізації запропонованих моделей і методів.

## ABSTRACT

Master's degree dissertation is devoted to solving the problem of methods efficiency increasing for evaluation of the automatic foreign texts translation systems.

As a result of the conducted researches the new model of language translation process was offered, the method of statistical evaluation of the automatic translation systems of language is improved, the procedure of the automated evaluation of the language machine translation systems on the basis of the offered methods is developed. Software for realization of the offered models and methods is developed.

## ВСТУП

Актуальність теми. Проблема машинного перекладу мови виникла в середині ХХ ст. практично одночасно з появою перших комп'ютерів в зв'язку з необхідністю автоматизації обробки користувачами великих обсягів інформації, представленої іншомовними текстами. З тих пір вона пройшла через декілька стадій розвитку, в результаті чого на сьогодні розроблено і використовується декілька десятків комерційних систем машинного перекладу мови.

Сучасний етап інформатизації суспільства характеризується появою великої кількості програмних продуктів, в тому числі і для машинного перекладу, тому для користувачів актуальною стає проблема вибору найбільш ефективного автоматизованого перекладача текстів. В зв'язку з цим постає нагальною задача розробки методів оцінювання якості роботи систем машинного перекладу (МП).

Традиційно оцінювання якості цих систем необхідно проводити за такими критеріями: якість роботи програмного забезпечення, зручність інтерфейсу користувача, лінгвістична характеристика, організаційне значення, цінова ефективність.

Оцінювання якості машинного перекладу за допомогою людини є повним, але дуже дорогим, воно може тривати місяці і не може багаторазово використовуватись для різних типів текстів та систем машинного перекладу мови. Явною стає проблема створення методів автоматичного оцінювання систем машинного перекладу, які були б швидкими, мовно-незалежними, які були б на одному рівні з людським оцінюванням, і мали б невисоку вартість розробки. Програмне забезпечення, основане на цих методах повинне використовуватись як автоматизований дублер кваліфікованих експертів з перекладу, а також замість них у випадках, коли є потреба у швидких або частих оцінюваннях якості МП.

Останнім часом проблема оцінювання МП досить широко розглядається в іноземних публікаціях, зокрема в працях [1-11]. Одним з найбільш відомих на сьогодні підходів, що використовуються для оцінки систем (МП), є порівняння кількості збігів N-грам в перекладеному і еталонному текстах [1]. Досить докладно ця тема розглядається в роботах К. Рарінені, на результатах досліджень проведених ним в

роботах [1, 8] базується переважна більшість сучасних публікацій, зокрема роботи Винокурова А.А. [12, 13].

Однак цей метод в більшій мірі підходить для систем статистичного машинного перекладу оскільки слабо враховує синтаксичні та семантичні аспекти мови[14].

Більшість методів використовують процедуру послівного оцінювання. Така процедура дає прийнятні за точністю результати тільки для текстів не флективних мов[15, 16].

Проблеми машинного перекладу розглядаються в працях [17-29] та електронних журналах [30-37].

Результати аналізу існуючих методів оцінки МП [38, 39] та аналіз ефективності роботи автоматичних перекладачів, свідчать про наявність невирішених задач в даній області і зокрема про актуальність проблеми розробки вдосконалених методів оцінювання ефективності систем автоматичного перекладу мови, які краще враховують лінгвістичну інформацію та знаходяться на одному рівні із людським оцінюванням.

Зв'язок роботи з науковими програмами, планами, темами.

Робота виконана згідно плану кафедральної науково-дослідної роботи № 52К4 “Методи прийняття рішень на основі принципів обчислювального інтелекту”.

Мета і задачі дослідження.

Метою дисертаційної роботи є підвищення ефективності методів оцінювання систем автоматичного перекладу.

Для досягнення вказаної мети в дисертації поставлені такі задачі:

- дослідити існуючі методи оцінювання ефективності систем автоматичного перекладу мови і на цій основі запропонувати нові, більш ефективні методи;

- вдосконалити метод статистичного оцінювання систем автоматичного перекладу мови;

- запропонувати нову модель процесу перекладу мови;

- розробити методіку автоматизованого оцінювання систем машинного перекладу мови на основі запропонованих методів;

- розробити програмне забезпечення для реалізації розроблених математичних моделей;



- запропонувати алгоритм порівняння систем автоматичного перекладу мови.

Об'єкт дослідження – процес автоматичного перекладу мови.

Предмет дослідження – методи та алгоритми перекладу тексту в задачах оцінки ефективності систем автоматичного перекладу мови.

Методи дослідження – кореляційний аналіз, математичне моделювання.

Наукова новизна отриманих результатів.

На основі виконаних теоретичних і експериментальних досліджень розроблено рекомендації з удосконалення методів оцінювання систем автоматичного перекладу мови й отримані такі нові результати:

1. Удосконалено метод статистичного оцінювання систем автоматичного перекладу мови, який відрізняється від існуючих тим, що враховує повторні появи N-грам у реченні та порівнює морфеми замість слів, що дає змогу більш точно оцінити якість перекладу.

2. Вперше запропоновано модель процесу перекладу мови, яка відрізняється тим, що розглядає його, як процес передачі повідомлення зашифрованого текстом вхідної мови по каналу з шумами, що дає змогу оцінити основний параметр якості перекладеного тексту – його інформативність.

Практичне значення одержаних результатів.

На основі запропонованих методів і моделей розроблено аналітичні, алгоритмічні і програмні засоби:

- методику автоматизованого оцінювання систем машинного перекладу мови на основі запропонованих методів, що дозволяє підвищити її адекватність шляхом врахування морфологічної і синтаксичної структури речень;

- програмне забезпечення для реалізації розроблених математичних моделей, що дозволяє використовувати його як автоматизованого дублера кваліфікованих експертів з перекладу, а також замість них у випадках, коли є потреба у швидких або частих оцінюваннях;

- запропоновано алгоритм порівняння систем автоматичного перекладу мови, що дозволяє розпізнавати закономірності та характеристики, притаманні тим чи іншим системам перекладу.

Результати роботи впроваджені в навчальний процес кафедри КСУ, а також використовуються в наукових дослідженнях, що проводяться в рамках держбюджетної тематики кафедральної науково-дослідної роботи № 52К4 “Методи прийняття рішень на основі принципів обчислювального інтелекту”.

Достовірність теоретичних положень магістерської кваліфікаційної роботи підтверджується строгістю постановки задач, коректним застосуванням математичних методів під час доведення наукових положень, строгим виведенням аналітичних співвідношень, порівнянням результатів, отриманих за допомогою розроблених у роботі методів, з відомими, та збіжністю результатів математичного моделювання з результатами, що отримані під час впровадження розроблених програмних і апаратних засобів.

Апробація результатів роботи. Основні положення й результати досліджень доповідалися й обговорювалися на форумі молодих науковців ВНТУ (м. Вінниця, 2019).

Публікації. За тематикою дослідження опублікована робота в репозиторії ВНТУ.

## 1 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ І ПІДХОДІВ ДО ВИРІШЕННЯ ЗАДАЧІ ОЦІНЮВАННЯ СИСТЕМ МАШИННОГО ПЕРЕКЛАДУ МОВИ

Активний розвиток глобальної мережі Інтернет та накопичення великої кількості інформації, представленої іншомовними текстами, зумовлює необхідність автоматизації її обробки [42-46]. Незважаючи на постійне концептуальне і програмне оновлення існуючих автоматичних перекладачів, якість їх роботи ще далека від якості роботи професійного перекладача-людини. Це свідчить про наявність невирішених задач в даній області і про необхідність прикладення подальших зусиль для їх розв'язання. Автоматичні методи для оцінювання різних аспектів якості МП таких, як адекватність, логічність і інформативність забезпечують альтернативи дорогому і тривалому процесу оцінки МП людиною.

В даному розділі проводиться аналіз стану проблеми оцінювання систем машинного перекладу мови з метою виявлення найбільш актуальних задач, розв'язання яких може привести до покращення якості автоматичного перекладу та методів його оцінювання.

### 1.1 Аналіз об'єкта дослідження

В найбільш загальному випадку під машинним перекладом розуміють передавання змісту висловлювань, сформульованих на одній мові, засобами іншої мови за допомогою комп'ютерної програми [17,18]. Вперше думку про можливість машинного перекладу висловив Чарльз Беббідж (1791-1871), який розробив у 1836-1848 роках проект цифрової аналітичної машини - механічного прототипу електронних цифрових обчислювальних машин, що з'явилися через 100 років. Фактично історія машинного перекладу починається з "Джорджтаунського експерименту". У січні 1954 р. відбулася перша публічна демонстрація машинного перекладу з російської мови на англійську, здійсненого на машині IBM-701. Повідомлення про цю подію було опубліковано в журналі *Computers and Automation*, 1954, № 2.

Перші роботи з машинного перекладу в Європі були виконані в тогочасному СРСР в Всесоюзному Інституті Наукової і Технічної Інформації під керівництвом Д.Ю. Панова і реалізовані на машині БЭСМ в 1955 році, а також в Математичному інституті АН СРСР під керівництвом Ляпунова А.А. і реалізовані на ЕОМ "Стрела". Ці роботи поклали початок першому поколінню систем машинного перекладу мови. Системи першого покоління базувались на алгоритмах послідовного перекладу "слово за словом", "фраза за фразою". Можливості таких систем визначалися доступними розмірами словників, що прямо залежали від обсягу пам'яті комп'ютера. Переклад тексту здійснювався окремими реченнями, змістовні зв'язки між ними ніяк не враховувались. Такі системи називаються системами прямого перекладу [47].

Перший період розвитку систем машинного перекладу закінчився в 1967 році з появою доповіді Комісії Національної Академії наук США. Виходячи з реальної ситуації з перекладами в Сполучених Штатах і з оцінки показників вартості різних способів перекладу. Комісія прийшла до висновку про нерентабельність застосування машинного перекладу в умовах США. Хоча комісія і рекомендувала проводити подальші дослідження в цій області, її рішення привели до різкого зниження інтересу до проблеми машинного перекладу, що істотно загальмувало його розвиток в цілому.

Другий період розвитку машинного перекладу почався з нового підйому робіт у цій області в середині 70-х років. В цей час з'явилась ціла низка практично діючих дослідницьких і комерційних систем. На користь відродження МП в цей період діяла низка об'єктивних факторів, таких, як з одного боку, зростання потреби в перекладах, зумовлене розширенням співробітництва різних країн, а з іншого, бурхливим розвитком можливостей комп'ютерної техніки. Вісімдесяті роки стали часом переходу до широкого практичного використання систем машинного перекладу в багатьох країнах, утворився ринок комерційних систем. Розроблено ряд глобальних проектів, таких як EUROTRA (європейське економічне співтовариство), ARIANE (Франція), METAL (США і Німеччина), KANT (США), SUSY (Німеччина), MU (Німеччина) [18]. Появився значний досвід експлуатації, обговоренню якого було присвячено багато конференцій. В цей час на зміну системам прямого перекладу прийшли системи наступних поколінь, в яких переклад тексту з мови на мову здійснюється на рівні синтаксичних структур. В алгоритмах перекладу використовувався набір операцій,

який дозволяв шляхом аналізу речення, що перекладалось, побудувати його синтаксичну структуру за правилами граматики вхідного речення, а потім перетворювати її в синтаксичну структуру вихідного речення і синтезувати вихідне речення шляхом підстановки потрібних слів із словника. Такі системи називаються Transfer- або T- системи (transfer – перетворення).

З початку 90-х років почався третій період розвитку технології машинного перекладу, який був ініційований черговим ростом інтересу до машинного перекладу, пов'язаним з бурхливим розвитком Інтернет і необхідністю online – перекладу[48]. В цей період продовжують розвивати започаткований в деяких попередніх проектах підхід до побудови систем машинного перекладу на основі отримання деякого, незалежного від мов, представлення змісту вхідного речення шляхом його семантичного аналізу (перехід до глибинної форми) з наступним синтезом після цього вихідного речення (перекладу) з отриманого смислового представлення. Такі системи МП називають I-системами ( аббревіатура від слова “Інтерлінгва”). Системи типу INTERLINGUA припускають апріорі наявність деякої метамови структур, на якій можна описати всі структури як вхідної, так і вихідної мов у загальному випадку; тому алгоритм перекладу в системі типу INTERLINGUA передбачається як більш простий: аналіз вхідного речення в термінах метамови і потім синтез з метаструктури відповідного речення вихідної мови. "Єдина" складність у цьому випадку - розробити саму метамову і описати природну мову у відповідних термінах. На сьогодні існує впевненість, що системи майбутніх поколінь будуть відноситись до класу I-систем.

## 1.2 Класифікація систем МП

Як уже було вказано, на разі існує велика кількість систем МП різного призначення. Оцінка якості їх роботи можлива за умови урахування їх цільового призначення. Згідно з цією характеристикою, системи МП можна поділити на такі класи [49-52].

1. Системи машинного перекладу (МП). Системи машинного перекладу - програми, що здійснюють цілком автоматизований переклад. Головним критерієм програми є якість перекладу. Крім цього, для користувача важливими моментами є зручність інтерфейсу, легкість інтеграції програми з іншими засобами обробки документів, вибір тематики, утиліта поповнення словника. З появою Internet основні

постачальники систем МП включили у свої продукти Web-інтерфейси, забезпечивши при цьому їхню інтеграцію з іншим програмним забезпеченням і електронною поштою, що дозволило застосовувати механізми МП для перекладу Web-сторінок, електронних листів і онлайн-розмовних сеансів.

2. Системи з функцією Translation Memory (TM). Надають кошти для так названого Machine Assisted Human Translation (МАНТ) - перекладу, виконуваного людиною за допомогою машини. Первісне призначення систем TM - полегшення роботи перекладачів при локалізації програмних продуктів і створенні термінологічних баз даних, надалі системи стали розвиватися як допоміжний інструментарій перекладача. Особлива увага приділяється підтримці форматів усіляких процесорів.

3. Систем з контрольованою мовою і машинний переклад на основі бази знань. У системах на основі контрольованої мови реалізований перехід від вільної вхідної мови до контрольованої вхідної мови. Контроль вхідної мови передбачає певні обмеження лексики, граматики, семантики. Контрольована вхідна мова використовується для спрощення виразів вихідного тексту, щоб підвищити якість перекладу.

4. On-line перекладачі. Служби онлайн-перекладу виконують переклад прямо у вікні браузера (mail.ru, google translate), не вимагаючи установки програми на комп'ютер користувача. Як правило, існує обмеження на обсяг тексту, що вводиться.

5. Словники on-line. Існує можливість пошуку, є рядок введення і меню для вибору вхідної і вихідної мови, серед словників - LOGOS ONLINE DICTIONARY (англ, фран, нім, ісп, італ, чеський), Dictor (100 мовних пар), TRADOS MultiTerm, Langenscheidt's New College Dictionary on-line і т.д.

6. Послуги по перекладу в Інтернеті. Практично кожна компанія - виробник систем МП пропонує послуги по перекладу. Наприклад, на сторінці Language Partners International організована служба перекладів, що обробляє будь-який електронний документ за допомогою систем МП і відправляє назад без редагування.

За способом організації систем перекладу мови їх можна класифікувати за наступними схемами [53,54]:

1.  $T_{вх} \Rightarrow МП \Rightarrow T_{мп} \Rightarrow Ред \Rightarrow T_{вих}$  ;

2.  $T_{\text{вх}} \Rightarrow \text{МП} \Rightarrow T_{\text{мп}} \Rightarrow \text{Ред} + \text{АРП} \Rightarrow T_{\text{вих}} ;$

3.  $T_{\text{вх}} \Rightarrow \text{Ред} + \text{АРП} \Rightarrow T_{\text{лп}} \Rightarrow \text{МП} \Rightarrow \text{Ред} + \text{АРП} \Rightarrow T_{\text{мп}} \Rightarrow T_{\text{вих}} ;$

4.  $T_{\text{вх}} \Rightarrow \text{ЛМП} \Rightarrow T_{\text{лмп}} \Rightarrow \text{Ред} + \text{АРП} \Rightarrow T_{\text{вих}} ;$

В схемах прийняті такі позначення:  $T_{\text{вх}}$  – вхідний текст, що підлягає перекладу;  $T_{\text{мп}}$  – перекладений машиною текст; МП – система машинного перекладу; ЛМП – система людино-машинного перекладу; Ред – професійний перекладач-редактор; АРП – засоби автоматизованої редакторської правки;  $T_{\text{лп}}$  – перекладений людиною текст;  $T_{\text{лмп}}$  – людино-машинний переклад тексту.

Схеми 1 і 2 – це схеми з наступним редагуванням перекладеного машиною тексту. Схема 2 на сьогодні має найбільше розповсюдження, в ній редагування здійснюється за допомогою засобів автоматизованого внесення редакторської правки, відомих під назвою word processing. Використання таких систем є ефективним з точки зору затрат часу і вартості за рахунок таких факторів. Відомо [19], що людина-перекладач витрачає 40% часу на пошук термінів в словнику, 40% на друкування перекладу і тільки 20% власне на переклад. Оскільки в схемі 2 скоротився час пошуку в словниках і на друкування тексту порівняно з перекладом, виконуваним людиною, то це і привело до більш швидкого виконання перекладу за даною схемою.

Корисною виявилась схема 3 перекладу з попереднім редагуванням, де перекладач-редактор працює з текстом ще до того, як його починає перекладати машина. В цьому випадку з вхідного тексту виключаються важкі для перекладу конструкції мови, або текст розмічається за допомогою дужок на прості речення.

В схемі 4 людина взаємодіє з системою машинного перекладу на проміжних стадіях, де вона бере на себе функцію подолання суттєвих труднощів (інтерактивний переклад). Такий переклад називають людино-машинним.

В подальшому, вживаючи термін “системи машинного перекладу”, будемо розуміти системи, що працюють за схемами 1-3.

### 1.3 Проблеми побудови систем МП

Основні проблеми МП витікають із низки труднощів, притаманних перекладу взагалі [55]. Перше джерело труднощів полягає в тому, що не можна перекласти текст

слово в слово: у мові перекладу може не існувати еквівалента для якогось слова оригіналу, крім того, різні мови майже завжди вимагають різного порядку слів; це означає, що послівний переклад буде невірний з погляду граматики. Під час перекладу складність представляють іноді навіть прості слова, не кажучи вже про переклад ідіоматичних слів і виразів у їхньому лінгвістичному контексті, щоб правильно передати зміст, що стоїть за ними, відповідно до правил, що встановлюються мовою перекладу. Інше джерело труднощів полягає в тім, що мови відрізняються одна від одної, іноді дуже сильно. У деяких мовах можуть взагалі не існувати засоби для передавання цілого ряду понять. Крім того, лінгвістичні засоби висловлення цих понять (що несуть при цьому і власне змістовне навантаження) на одній мові можуть не існувати в інших мовах, чи використовуватися в них для інших цілей. Отже, під час перекладу доводиться нерідко передавати далекі поняття, перебудовуючи при цьому лінгвістичні структури, використовувані для передавання цих понять, намагаючись одночасно звести до мінімуму втрати інформації. Технічний переклад тільки з однієї мови вимагає звертання до об'ємних словників (20 - 50 тисяч термінів на кожну область техніки), не враховуючи 10 - 20 тисяч загальноживаних слів, з якими, як правило, доводиться мати справу під час перекладу [20]. Крім того, потрібне постійне оновлення словників, особливо технічних, що зумовлено постійним розвитком техніки і, відповідно, виникненням нових термінів і слів, що їх позначають.

Ще одним джерелом труднощів МП є відсутність потужних баз знань, які б моделювали достатньо точно навколишній реальний світ. Розроблені на сьогодні в рамках штучного інтелекту небагаточисельні моделі реального світу є досить обмеженими як за широтою, так і за глибиною охоплення. Це не дозволяє перевіряти можливі інтерпретації вхідних речень, усувати неоднозначності, і, головне, генерувати переклад. Потрібно також відзначити, що сучасні методи представлення знань не дозволяють в принципі моделювати в повній мірі реальний світ чи, хоча б, якусь його значиму частину.

Крім того, не має на сьогодні і підходящих лінгвістичних методів. Природні мови досить неоднозначні, і це не дає можливості будувати на їх аналіз тексту тільки на їх синтаксисі [56-58]. Через це необхідні семантичні засоби контролю, які могли б в принципі дозволити системі МП функціонувати, опираючись перш за все на



синтаксис, на сьогодні не розроблені достатньо чітко для їх універсального використання. Тому системи МП можуть сьогодні успішно застосовуватись тільки до визначених, зарані відомих мовних явищ.

Таким чином, як висновок можна зазначити, що основними проблемами про розробці систем МП є такі:

- проблема створення великих словників для систем;
- проблема навчити систему розпізнаванню стійких мовних оборотів, такі, наприклад, як “ПРИВІТ, ЯК СПРАВИ?”;
- проблема записати всі правила перекладу речень у вигляді програми.

Розглянемо, наприклад, способи розв’язання цих проблем у системах машинного перекладу, розроблених російською фірмою “PROMT”, які, на сьогодні, за світовим визнанням, мають найкращі показники в роботі.

У системах сімейства PROMT [59,61] розроблено практично унікальний за повнотою морфологічний опис для всіх мов, з якими системи вміють працювати. Він містить 800 типів словозміни для російської мови, більш 300 типів, як для німецької, так і для французької мови, і навіть для англійської, котра не належить до флективних мов, виділено більш 250 типів словозміни. Безліч закінчень для кожної мови зберігається у вигляді деревних структур, що забезпечує не тільки ефективний спосіб збереження, але й ефективний алгоритм морфологічного аналізу. Крім того, використовувана модель морфології дозволила розробити експертну систему для користувача-розробника словника. Ця система фактично автоматизує процедуру виділення основи і визначення типу словозміни при введенні нових слів.

Глобальні проекти з розробки систем МП були орієнтовані на розв’язання задачі перекладу в цілому. У них розробка опису лексичних одиниць у словнику і розробка алгоритмів перекладу розглядалися як різні задачі. З’явилася безліч лінгвістичних робіт, що пропонували структуру опису властивостей живого слова в словниковій статті машинного словника.

Локальні проекти були орієнтовані на рішення тільки вузьких задач. Установкою розроблювачів було одержання хоч якихось практичних результатів. У них словниковий опис і опис алгоритмів розглядалися як сторони однієї проблеми, але рішення, як правило, шукалося в обмеженні розглянутого світу, або граматичного, або

семантичного. Наприклад, на основі ознаки "приналежність до частини мови" описувалася граматики такого типу:

іменна група - це іменник

іменна група - це прикметник + іменна група

дієслівна група - це дієслово + іменна група

речення - це іменна група + дієслівна група

Зрозуміло, що деяка частина речень природної мови описується такою граматикою, але ця частина дуже незначна, і на її основі не можна правильно аналізувати і переводити хоч скільки-небудь реальний текст. Але зате можна використовувати ефективні методи побудови перетворювача по заданій чи граматиці, на худий кінець, написати програму, що шляхом перебору побудує дерева залежностей для обмеженої безлічі речень [62].

Однак при розробці систем PROMT уперше був застосований фактично революційний підхід, що і дозволив одержати вражаючі результати [61]. Системи перекладу сімейства PROMT - це системи, спроектовані на основі не лінгвістичних, а кібернетичних методів.

Виявилось, що дуже продуктивно розглядати систему перекладу не як транслятор, задачею якого є переклад тексту, припустимого з погляду вхідної граматики, а як деяку складну систему, задачею якої є одержання результату при довільних вхідних даних, у тому числі і для текстів, що не є правильними для граматики, з яким працює система. Замість прийнятого лінгвістичного підходу, що припускає виділення послідовних процесів аналізу і синтезу речення, в основу архітектури систем було покладене представлення процесу перекладу як процесу з "об'єктно-орієнтованою" організацією, заснованої на ієрархії оброблюваних компонентів речення [58]. Це дозволило зробити системи PROMT стійкими і відкритими. Крім того, такий підхід дав можливість застосування різних формалізмів для опису перекладу різних рівнів. У системах працюють і мережні граматики, близькі по типі до розширених мереж переходів, і процедурні алгоритми заповнення і трансформацій фреймових структур для аналізу складних предикатів.

Система PROMT виконує переклад типу TRANSFER. Але це дуже проста відповідь, вона практично не відбиває особливостей архітектури системи PROMT. А

особливості полягають у тому, що метод TRANSFER застосований у системі не у відповідності зі стандартним лінгвістичним підходом.

Оскільки система МП, як правило, працює в умовах не цілком описаних даних, адже в мові, що розвивається дуже швидко постійно з'являються нові слова, нові функції старих слів, і, разом з новими сутностями, нові значення. У цих умовах визначальним структурною властивістю алгоритмів перекладу стає їхня стійкість до довільних вхідних даних, і в основу алгоритмів, що виконують переклад у системі PROMT, замість послідовного TRANSFER був закладений ієрархічний підхід, що розділяє процес перекладу на взаємозалежні TRANSFER для різних одиниць аналізу.

У системі виділяється рівень лексичних одиниць, рівень груп, рівень простих речень і рівень складних речень. Усі ці процеси зв'язані і взаємодіють ієрархічно відповідно до ієрархії текстових одиниць, обмінюючи синтезованими і наслідуваними ознаками [56].

Аналіз простих речень як структур, що складаються із синтаксичних одиниць, виконується на основі фреймових предикатних структур, що дозволяють ефективно виконувати перетворення. Дієслово вважається для простих речень головним елементом і його валентності визначають заповнення відповідного фрейму. У такий спосіб здійснюється TRANSFER на рівні речень. Під часу аналізу складних речень виконується узгодження часів і правильний переклад сполучників.

Крім розвитку власне технології машинного перекладу були прикладені спеціальні зусилля до того, щоб реалізувати в нових системах можливість лінгвістичного настроювання, що забезпечує значне поліпшення якості перекладу для спеціалізованих текстів. Зокрема, появу підтримки тематик у новому ядрі дозволило кардинально змінити підхід до перекладу спеціалізованих текстів [51]. Використання тематик вирішило проблему лінгвістичного настроювання системи на спеціалізовані тексти і забезпечило ефективність їхнього збереження. Легкість і зручність лінгвістичного адміністрування систем МП було реалізовано на базі наступних технологій:

*AutoDetect* для реалізації автоматичного визначення теми документа,

*SmartFolders* для попереднього завдання тематики перекладу всім документам, що знаходяться у визначеній папці,

*KeyWords* для вибору тематики на основі попереднього заданого списку ключових слів

*SmartURL* для забезпечення можливості попереднього завдання тематики перекладу для конкретного URL.

#### 1.4. Аналіз існуючих принципів функціонування та методів оцінювання систем МП

В наш час існують основні три принципи, на яких будуються системи МП [17,18]:

1. Системи типу Transfer (перетворення) – в алгоритмі перекладу використовується набір операцій, який дозволяє шляхом аналізу речення, що перекладається, побудувати його синтаксичну структуру за правилами граматики вхідного речення, а потім перетворювати її в синтаксичну структуру вихідного речення і синтезувати вихідне речення шляхом підстановки потрібних слів із словника.

2. Системи типу INTERLINGUA – алгоритм побудови припускає наявність деякої метамови структур, на якій можна описати всі структури як вхідної, так і вихідної мов у загальному випадку; тому алгоритм перекладу в системі типу INTERLINGUA передбачається як більш простий: аналіз вхідного речення в термінах метамови і потім синтез з метаструктури відповідного речення вихідної мови.

3. Системи статистичного декодування - характеризується повторенням на більш високому рівні принципу прямого перекладу, що використовувався в перших системах МП. Цей підхід полягає в використанні відповідностей між так званими N-грамами [1] вхідного і вихідного текстів для прямого перекодування першого в другий. Ці відповідності знаходяться статистичними методами шляхом навчання на вибірках масивів вхідного і правильно перекладеного тексту [8], при цьому лінгвістичні характеристики текстів не використовуються.

##### 1.4.1. Метод оцінки якості перекладу на основі N-грам

Системи статистичного декодування обігнали колись привілейовану, традиційну технологію, яку лише частково можна назвати автоматичною. В основі її лежать

багатомовні словники, які програмісти й перекладачі збирають вручну, і на основі яких згодом "навчають" програму, яке слово що означає - з урахуванням синтаксису.

Нова технологія статистичного декодування дозволяє оцінювати системи, які будуть перекладати тексти на самих різних мовах, головним є наявність "паралельних даних".

Основна ідея системи статистичного декодування дуже проста: якщо комп'ютеру надати текст англійською мовою, і виконаний вручну його переклад на іншу мову, то машина, використовуючи статистичний метод, "вивчить" другу мову[1,4].

Наприклад порівнюються дві прості фрази на арабському (написане латиною) - "raj1 kabir" і "raj1 tawil". Якщо машина "знає", що перша фраза означає "велика людина", а друга - "висока людина", то проста статистика дозволить машині зрозуміти, що "raj1" означає "людина". Подібні прості фрази зветься "N-грами", і вони вважаються базовими елементами систем статистичного декодування.

Одним з найбільш відомих на сьогодні підходів, що використовуються для оцінки систем (МП) є порівняння кількості збігів N-грам в перекладеному і еталонному текстах. Хоча принцип здається простим до примітивізму, у нього є ряд своїх "але". По-перше, для повноцінного порівняння потрібен величезний масив даних - одним текстом не обійдешся. По-друге, для ефективної роботи таких систем потрібні комп'ютерні потужності, які ще на початку 1990-х були ще недоступні.

Як вимірюється якість перекладу?

Чим ближче машинний переклад до професійного людського перекладу, тим він кращий[1-11, 24-26]. Це і є центральною ідеєю оцінки якості систем машинного перекладу мови. Щоб оцінити якість машинного перекладу, вимірюється його близькість до одного або декількох перекладів людини. Таким чином, наша система оцінки МП вимагає двох компонентів:

- а) числова близькість перекладу.
- б) набір якісних перекладів за допомогою людини-експерта.

Ми вираховуємо близькість, порівнюючи машинний переклад та набір якісних перекладів за допомогою людини-експерта й обраховуємо кількість розходжень при

виборі слова й при виборі порядку слів.

Як правило, є багато якісних перекладів даного вихідного речення. Ці переклади можуть відрізнитись вибором слова або порядком слів, навіть коли вони використовують ті ж самі слова. І все-таки люди можуть якісно відрізнити гарний переклад від поганого. Наприклад, розглянемо ці два машинних переклади (МП) китайського вихідного речення:

Приклад 1:

Переклад МП 1: It is a guide to action which ensures that the military always obeys the commands of the party.

Переклад МП 2: It is to insure the troops forever hearing the activity guidebook that party direct.

Хоча вони на перший погляд здається, описують одне і теж, але вони помітно відрізняються по якості. Для порівняння, ми пропонуємо три переклади людини-експерта (ПЕ) того ж самого речення нижче:

Приклад ПЕ 1: It is a guide to action that ensures that the military will forever heed Party commands.

Приклад ПЕ 2: It is the guiding principle which guarantees the military forces always being under the command of the Party.

Приклад ПЕ 3: It is the practical guide for the army always to heed the directions of the party.

Ясно, що кращий машинний переклад 1, тому, що він має багато слів і фраз спільних із цими трьома людськими перекладами, у той час як переклад МП 2 не містить цього.

Відмітимо, що МП 1 має співпадіння фрази " It is a guide to action " з МЕ 1, " which " з МЕ 2, " ensures that the military " з МЕ 1, " always " з МЕ 2 і 3, " commands " з МЕ 1, і нарешті " of the party " з МЕ 2. А МП 2 має набагато менше співпадінь і їхній ступінь – менший.

Ясно, що програма може оцінити МП 1 вище чим МП 2 просто, порівнюючи кількості співпадінь N-грама у кожному машинному перекладі й перекладі за допомогою експертів. Експерименти показують, що ця властивість оцінювання - загальне явище, а не відноситься тільки до декількох простих прикладів [1, 8]. Первинне завдання для створення програми порівняння полягає у тому, щоб зрівняти N-грами МП з N-грамами ЕП і порахувати число співпадінь. Ці співпадінь не повинні залежати від положення

слова в реченні. Чим більше співпадіннь, тим кращий машинний переклад.

#### 1.4.2 Метод оцінки якості перекладу на основі зваженої моделі N-грам

Серйозна проблема для методу BLEU - це необхідність великої кількості варіантів перекладів людиною. BLEU схильний вважати помилкою будь-яке відхилення від базового набору N-грам, не дивлячись на те, що звичайно існує декілька прийнятних варіантів перекладу однієї і тієї ж частини тексту, і тому будь-які відхилення в якості МП можуть бути тільки частково віднесені до помилок. Звичайно, потрібно багато різних варіантів перекладу експертів для одного і того ж тексту для того, щоб вирішити цю проблему, але це робить автоматичну оцінку більш дорогою.

Пропонується результат експерименту по розробці альтернативного вирішення цієї проблеми: домноження N-грам порівнянь на вагові коефіцієнти, які характеризують потенційну стабільність N-грам в межах різних варіантів ручних перекладів[63, 64].

Ця ідея виникла з області відбору інформації: деякі фрагменти інформації в тексті мають більш важливе - ключове значення (наприклад, імена подій і учасників події), і тому мають зберігатись в будь-якому перекладі, менш важлива інформація (наприклад, вибір морфологічно-синтаксичного контексту і певні функціональні слова) підлягає більшій варіації, оскільки перекладач має більшу свободу вибору.

Вага для N-грам автоматично генерується на підставі співставлення їх частоти в даному тексті і в сукупності текстів, які потрібно перекласти. Отримане таким чином ранжування N-грам тестується за допомогою їх порівняння із дійсною стабільністю в двох доступних ручних перекладах для 100 текстів системі оцінювання МП DARPA-94. Результуючі оцінки BLEU корегуються вагою стабільності N-грам: вклад зазначених N-грам в оцінку МП пропорційний до призначеної ваги.

Запропонований підхід до оцінки МПА перевірений на наборі перекладів із французької мови на англійську різними системами МП, доступними в системі DARPA, і звірений з результатами базового методу BLEU.

Оцінки, отримані на основі моделі зважених N-грам, повинні бути узгоджені з базовими результатами оцінки BLEU для логічності і перевершувати оцінки BLEU в

адекватності. Крім того, вони можуть бути також надійно використовуватись, якщо є тільки один ручний переклад для тексту, який оцінюється.

Окрім збереження коштів, можливість надійно використовувати єдиний ручний переклад має додаткову перевагу: стає можливим створення методів оцінки МП, основаних на вибірці, що було проблематично для оцінки з використанням багатьох ручних перекладів, так як тільки один обраний варіант із набору використовується в перекладі (Parinen і інші. 2001:4). Проте, вибірка зважених N-грам, як виявляється, дає якісну оцінку суджень людини про адекватність перекладу. Використання зважених N-грам дуже важливе для прогнозу адекватності, оскільки кореляція вибірки для незважених N-грам набагато нижча.

Слова в тексті несуть унікальне інформаційне навантаження [65], і в результаті мають різну важливість для перекладу. Більш інформативні (наприклад, контент) слова мають схильність повторюватися у різних ручних перекладах. Точна обробка (передача) цих слів системою МП вказує на якість перекладу. Проте, автоматичні методи, які використовують ручний переклад як основу, неявно припускають, що всі слова однаково важливі в ручному перекладі, і використовують їх усі в підрахунках N-грам [1, 8] або у вимірюванні меж редагування [2, 3]. Змінна частина різних ручних перекладів є, в багатьох випадках, обмеженою вибором синтаксичного контексту для стійкої основи наявних слів. Таким чином, отримуються більш правильні результати оцінки МП навіть при використанні одного ручного перекладу, якщо оцінки зазначених N-грам скоректовані з вагою, яка відображає значення слова в даному тексті.

Модель зважених N-грам виконана як програмний пакет для оцінки МП (який включає скрипт Perl, файли прикладів і документацію), доступна для вільного використання за електронною адресою [66].

В експерименті використовувались французько - англійські переклади, доступні в системі оцінки МПА DARPA-94. Система містить 100 французьких текстів, перекладених на англійську 4 різними системами МП: “Systran”, “Globalink”, “Метал” і “Candide”. Вона також містить 2 незалежних ручних переклади кожного тексту і оціночні бали, даних людьми-експертами кожному із 400 машинних перекладів.



Оцінки експертів були дані для 3 параметрів якості перекладу: “адекватність”, “логічність” і “Інформативність”.

На першому етапі експерименту, один набір ручних перекладів використовувався, щоб обчислити “значимість” S-бал для кожного слова в кожному з 100 текстів, як пропонується в [6]. Табл. 1 описує процедуру обчислення S-балів.

В другій стадії, яку було виконано - оцінка МП на основі N-грам, вимірювалась точність і вибірка N-грами вихідного МП, що використовувала єдиний ручний переклад. Підрахунки N-грам були скоректовані за допомогою S-балів для кожного зазначеного слова.

Зважені бали точності і вибірки були перевірені на корельованість оцінкам людини для тих же текстів і порівняні з результатами подібних тестів для стандартної оцінки BLEU. Щодо оцінювання систем МП, то кореляція для моделі зважених N-грам виявилась кращою, як для адекватності так і логічності, запропоноване удосконалення методу виявилось кращим для оцінювання адекватності. Ці результати свідчать, що модель зважених N-грам дає набагато точніші прогнози про статистичну систему МП “Candide”, тоді як стандартний підхід BLEU прагне переоцінити її продуктивність.

Значимість S-балів основана на співставленні відносних частот слів в специфічному тексті і в решті сукупності ручних перекладів. Вони обчислені за формулою [65]:

$$S_{word[text]} = \ln \frac{(P_{word[text]} - P_{word[rest-corp]}) \times N_{word[txts-not-found]}}{P_{word[all-corp]}} \quad (1.1)$$

де  $S_{word[text]}$  - це оцінка статистичної значимості для певного слова в певному тексті;

$P_{word[text]}$  - є відносною частотою слова в тексті;

$P_{word[rest-corp]}$  - є відносною частотою того ж слова в решті текстів (сукупності текстів), без цього тексту;

$N_{word[txts-not-found]}$  - відносна частина текстів у сукупності, де це слово не знайдене (число текстів, де це слово не знайдене поділене на число текстів в сукупності);

$P_{word[all-corp]}$  - є відносною частотою слова в усій сукупності текстів.

Перший коефіцієнт ( $P_{word[text]} - P_{word[rest-corp]}$ ) в цій формулі - це різниця відносних частот в певному тексті і в решті текстів із сукупності. Його значення дуже високе для правильних іменників, які повторно з'являються в одному тексті, але мають дуже низьку (часто 0) частоту в решті текстів із сукупності. Чим вища різниця, тим більшою є значимість слова для цього тексту [64].

Другий коефіцієнт  $N_{word[txts-not-found]}$  описує, наскільки рівномірно слово розповсюджується по сукупності текстів: якщо це поширене в невеликій кількості текстів, оцінка висока і слово має більше шансів стати статистично значимим для певного тексту.

Третій коефіцієнт ( $1 / P_{word[all-corp]}$ ) впорядковує слова за їх відносними частотами, таким чином, слова із низькою частотою появи і слова із високою частотою появи у мові мають рівні шанси стати статистично значимими. Передбачається, що автор тексту має повний “умоглядний контроль” над значимістю слів.

Натуральний логарифм обчисленої оцінки використовується, щоб поступово знизити ряд оціночних S-балів.

Гранична величина для відділення смислових слів і функціональних слів була знайдена за допомогою експерименту:

$S-score = 1$ ;  $S-score < 1$  для (більшості із) функціональних слів;  $S-score > 1$  для смислових слів.

Цей поріг був знайдений, для того щоб відрізнити смислові слова і функціональні слова також в інших окрім англійської мовах, це було перевірено для французької мови і російської мови, що дало подібні результати.

Таблиця 1.1 - Результати впорядкування слів на прикладі тексту 1. згідно їх S-балів

| r | S     | word         | r  | S      | word      |
|---|-------|--------------|----|--------|-----------|
| 1 | 2,918 | OPEC         | 8  | 1,844  | total     |
| 1 | 2,918 | Emirates     | 8  | 1,844  | report    |
| 1 | 2,918 | barrels      | 9  | 1,692  | current   |
| 1 | 2,918 | oil          | 10 | 1,593  | price     |
| 1 | 2,918 | quota        | 10 | 1,593  | news      |
| 1 | 2,918 | Subroto      | 11 | 1,470  | recent    |
| 1 | 2,918 | world        | 12 | 1,270  | month     |
| 1 | 2,918 | cartel       | 13 | 1,161  | officials |
| 1 | 2,918 | war          | 14 | 0,972  | because   |
| 1 | 2,918 | ruler        | 15 | 0,805  | million   |
| 1 | 2,918 | petroleum    | 16 | 0,781  | yesterday |
| 1 | 2,918 | markets      | 17 | 0,651  | That      |
| 1 | 2,918 | gestures     | 18 | 0,621  | also      |
| 1 | 2,918 | estimates    | 19 | 0,527  | much      |
| 1 | 2,918 | conciliatory | 20 | 0,331  | But       |
| 1 | 2,918 | Zayed        | 21 | 0,291  | over      |
| 1 | 2,918 | UAE          | 22 | 0,007  | from      |
| 1 | 2,918 | Szabo        | 23 | -0,079 | there     |
| 1 | 2,918 | Sheik        | 24 | -0,126 | after     |
| 1 | 2,918 | Saudi        | 25 | -0,233 | their     |
| 1 | 2,918 | Petroleum    | 26 | -0,244 | new       |
| 1 | 2,918 | Dhabi        | 27 | -0,284 | had       |
| 1 | 2,918 | Arabia       | 28 | -0,411 | as        |
| 1 | 2,918 | Abu          | 29 | -1,225 | talks     |
| 2 | 2,719 | output       | 30 | -1,388 | been      |

Продовження таблиці 1.1

| r | S     | word       | r  | S      | word |
|---|-------|------------|----|--------|------|
| 3 | 2,449 | others     | 31 | -1,594 | at   |
| 3 | 2,449 | manager    | 33 | -1,844 | on   |
| 3 | 2,449 | government | 34 | -2,214 | its  |
| 3 | 2,449 | dropped    | 35 | -3,411 | for  |
| 3 | 2,449 | declines   | 36 | -3,707 | with |
| 3 | 2,449 | agency     | 38 | -4,238 | The  |
| 4 | 2,375 | day        | 39 | -4,319 | by   |
| 5 | 2,305 | production | 40 | -4,458 | Mr   |
| 6 | 2,096 | well       | 41 | -5,323 | the  |
| 6 | 2,096 | demand     | 42 | -      | a    |
| 7 | 1,880 | concern    | 42 | -      | of   |

Текст 1: “[...] After recent major price declines because of a sharp rise in OPEC production, total cartel output has now dropped as much as one million barrels a day -- from

the year's peak of 20 million barrels a day early this month. And there are new indications that even the United Arab Emirates, OPEC's most notable quota violator, might soon cut its own output below a targeted 1.5 million barrels a day because of pressure from others in the cartel, including Saudi Arabia.

The Emirates aren't making any promises publicly. But its government did offer conciliatory gestures to OPEC yesterday. [...]

Припущення, що ранжування за S-балами представляє центрованість даних понять для цього тексту, може бути корисним для цілого ряду застосувань, включаючи автоматичну оцінку МП.

S-бали генерувалися для кожного слова кожного тексті в обох наборах ручних перекладів, доступних в системі DARPA [65].

В другій стадії експерименту S-бали значимості використовуються, як вага для встановлення коефіцієнтів відповідних N-грам, коли вихідні дані певної системи МП порівнюються з ручним перекладом.

Постає питання чи запропонований метод оцінки МП дозволяє використовувати єдиний ручний переклад надійно. Для того, щоб оцінити стабільність зважених оціночних балів маючи один ручний переклад, було виконано два ходи експерименту. Перший хід використовував “ручний переклад”, поки другий використовував ручний переклад експерта. Але кожного разу використовувався тільки один переклад. Оцінки для обох ходів були порівняні, використовуючи середньо квадратичне відхилення.

Щоб перейти від S-балів до оцінок N-грам використовувалася наступна процедура:

- Якщо для лексичного елемента в тексті  $S\text{-score} > 1$ , всі лічильники для N-грам, що містять цей елемент збільшуються на S-бал (не тільки на 1, як в базовому підході BLEU).
- Якщо  $S\text{-score} \leq 1$ ; застосовується звичайний підрахунок N-грам: лічильник збільшується на 1.

Розраховуються як звичайні оцінки, що використовуються для BLEU, так і зважені оцінки. Наступний код було додано до скрипту Perl інструменту BLEU: окрім оператора, який збільшує лічильник для кожного N-грам  $\$ngr$  на 1:

```

$ngr .= $words[$i+$j] . " ";
$$hashNgr{$ngr}++;
[.]
$WORD = $words[$i+$j];
$WEIGHT = 1;
if(exists
$WordWeight{$TxtN} {$WORD} &&
$WordWeight{$TxtN} {$WORD} >1){
$WEIGHT=
$WordWeight{$TxtN} {$WORD};
}
$ngr .= $words[$i+$j] . " ";
$$hashNgr{$ngr}++;
$$hashNgrWEIGHTED{$ngr}+= $WEIGHT;
[.],

```

де структура даних хеш  $\$WordWeight\{\$TxtN\}\{\$WORD\}=\$WEIGHT$  зображає таблицю S-балів для слів в кожному тексті корпусу, подібно Табл. 1.1.

S-бали роблять важливий внесок у вагу всіх N-грам в тестованих текстах, ручних перекладах, і для підрахунків N-грам, знайдених в обох наборах текстів.

Табл. 1.2 підсумовує цей вклад для зазначених N-грам. Це дало загальне представлення того, наскільки запропонований підхід залежить від значення ваги. Додана вага до підрахунків N-грам в перевіреному і в довідкових перекладах, представляє близько 97%–98% загальних балів, що використовуються для оцінки МП.

Таблиця 1.2 - Відповідність N-грам та S-балів

| Система МП | Кількість N-грам | Сума ваг S-балів | Додана вага (%) |
|------------|------------------|------------------|-----------------|
| candied    | 45074            | 1,654,396.0      | 97.2 %          |
| globalink  | 41700            | 1,594,201.5      | 97.4 %          |
| ms         | 44433            | 1,682,107.0      | 97.3 %          |
| reverse    | 46403            | 1,762,911.3      | 97.4 %          |
| systran    | 47102            | 1,799,162.3      | 97.4 %          |

Оціночні бали зважених N-грам для точності, Recall (повторної появи слова) і F-бали можуть рахуватись для частини, для цілого тексту або для сукупності перекладів, що генеруються системою МП.

В табл. 1.3 вказано наступні бали:

– людські оціночні бали для адекватності і логічності (середні відмітки для всіх текстів, генерованих кожною системою МП);

– бали BLEU, отримані з використанням 2 ручних перекладів (Розмір N-грам = 4);

– Точність, Recall і F-бали для моделі зважених N-грам використовуючи 1 ручний переклад і розмір N-грам = 4.

– Коефіцієнт кореляції Пірсона  $r$  для Точності, Recall і F-балів, який співвіднесений з людськими балами для адекватності і логічності. Перший ряд в кожному випадку показує кореляцію  $r(2)$  (з 2 ступенями свободи) для сукупності, що включає тільки бали для систем МП, але не “ ручного перекладу експерта” ; другий ряд показує кореляцію  $r(3)$  (з 3 ступенями свободи), коли до сукупності було додано ручний переклад.

Бали вверху кожної комірки показують результати для першого ходу експерименту, який використовував “Довідковий ручний переклад” ;бали внизу комірки показують результати для другого ходу з “ ручним перекладом експерта” .

Таблиця 1.3. - Оціночні бали для першого та другого ходів експерименту

| System<br>[адекв.] / [логічн.] | BLEU<br>[1&2] | Точність<br>(w) 1/2 | Recall<br>(w) 1/2 | F-бал<br>(w) 1/2 |
|--------------------------------|---------------|---------------------|-------------------|------------------|
| HT-Expert<br>0.921 / 0.852     |               | 0.7945              | 0.6685            | 0.7261           |
| CANDIDE<br>0.677 / 0.455       | 0.3561        | 0.6996<br>0.7020    | 0.5067<br>0.5072  | 0.5877<br>0.5889 |
| GLOBALINK<br>0.710 / 0.381     | 0.3199        | 0.6306<br>0.6318    | 0.4883<br>0.4876  | 0.5504<br>0.5504 |
| MS<br>0.718 / 0.382            | 0.3003        | 0.6217<br>0.6201    | 0.5152<br>0.5111  | 0.5635<br>0.5603 |
| REVERSO<br>NA / NA             | 0.3823        | 0.6793<br>0.6805    | 0.5400<br>0.5389  | 0.6017<br>0.6015 |
| SYSTRAN<br>0.789 / 0.508       | 0.4002        | 0.6850<br>0.6869    | 0.5511<br>0.5507  | 0.6108<br>0.6113 |
| Корел. $r(2)$ із [адекв] – МТ  | 0.5918        | 0.0726<br>0.0708    | 0.8347<br>0.8271  | 0.5686<br>0.5469 |
| Корел. $r(3)$ із [адекв]/МТ&НТ |               | 0.8080              | 0.9718            | 0.9355           |
| Корел. $r(2)$ із [лог] – МТ    | 0.9807        | 0.8641<br>0.8618    | 0.8017<br>0.8440  | 0.9802<br>0.9894 |
| Корел. $r(3)$ із [лог] МТ&НТ   |               | 0.9556              | 0.9819            | 0.9965           |

З табл. 1.3 видно, що є добра позитивна кореляція між базовими оцінками

BLEU і оцінками людини для логічності:  $r(2)=0.9807$ ,  $p < 0.05$ . Проте, кореляція з адекватністю набагато гірша і статистично не значна:  $r(2)= 0.9807$ ,  $p > 0.05$ . Найсерйозніша проблема для BLEU – бали прогнозування для статистичної системи МП “Candide”, переклад якої було оцінено як відносно логічний, але в значній мірі неадекватний. Для інших систем МП бали для адекватності і логічності узгоджені одні з одним: більш логічні переклади також є адекватним. Бали BLEU співпадають з балами логічності “Candide’s”, не рахуючи бали Адекватності. Коли “Candide” виключався з сукупності кореляція  $r$  підвищувалась, але все була нижчою, ніж кореляція із логічністю і залишається статистично незначимою:  $r(1)=0.9608$ ,  $p > 0.05$ . Таким чином, базовий підхід BLEU не в змозі послідовно передбачити бали для адекватності.

Запропонована модель зважених N-грам перевершує BLEU в здатності передбачити бали для адекватності: зважені бали Recall мають набагато більш сильну кореляцію з відповідністю (але, все ще, залишаються статистично незначимими для оцінки МП:  $r(2)=0.8347$ ,  $p > 0.05$ ;  $r(2)= 0.8271$ ,  $p > 0.05$ , але які стають значимими, коли бали для ручних перекладів додаються до сукупності:  $r(3)= 0.9718$ ,  $p < 0.01$ ).

Це досягається за рахунок зменшення переоцінки системи “Candide”, наближаючи її бали оцінки, ближче до людських суджень про якість в даному відношенні. Проте, це повністю не досягнуто: “Candide” все ще є трохи кращою, ніж Globalink згідно зваженим балам Recall, але є трохи гіршою, ніж Globalink згідно людських суджень про адекватність.

Для обох методів – BLEU і оцінки на основі зважених N-грам – адекватність прогнозувати важче, ніж логічність [1,8,56]. Це завдяки тому, що не існує лінгвістичної моделі відповідності перекладу, яка може бути легко формалізованою. Введення зважених S-балів може бути корисним кроком у напрямку до розробки такої моделі, так як бали кореляції з Відповідністю набагато кращі у підході зважених N-грам, ніж для BLEU.

По-перше припускається, що довідковий переклад є адекватним оригіналу і дійсно, оцінює відповідність “одномовного перекладу” поганою англійського” “нормальному англійському”. Проте, люди, які оцінюють відповідність двомовних перекладів використовують різні набори даних для їх оцінювання: чи є чи ні ручний

довідковий переклад адекватним оригіналу сумнівно. Звичайно, базовий тут не є 1.0 бал: відзначу, що альтернативний ручний переклад (який я використовую як довідковий в другому ході експерименту) мав не 1.0 бал, але 0.921 і 0.852 для адекватності та логічності відповідно, згідно людських суджень.

Також із лінгвістичної точки зору, вага S-балів і N-грам може бути виправдано хорошим наближенням адекватності, яка залучає широкий ряд коефіцієнтів, таких, що стосуються синтаксичних і семантичних питань, які неможливо охопити оцінками N-грам, і вимагають тезаурус і інші додаткові бази знань. Точні формальні моделі перекладу можуть бути також корисними для поліпшення автоматичної оцінки Відповідності.

Запропонований метод оцінки також забезпечує здатність BLEU послідовно прогнозувати бали для логічності. Зважені F-бали мають сильну позитивну кореляцію з цим аспектом якості МП, числа дуже подібні значенням для BLEU:  $r(2)=0.9802$ ,  $p<0.05$ ;  $r(2)=0.9893$ ,  $p<0.01$ ;  $r(3)=0.9965$ ,  $p<0.01$ .

Проте, сильна кореляція з адекватністю в запропонованому методі досягається іншими засобами, ніж в BLEU: замість “точності N-грам”, що пропонується в (Papineni, 2001:2) і 2 ручних довідкових переклади, було використано об'єднаний F-бал (Точність і Recall, зважені однаково) і тільки 1 ручний довідковий переклад. Підрахунки зважених N-грам були прості; ніякі модифікації стандартних вимірювань точності і Recall не застосовувались.

Ці дві головні методологічні відмінності пов'язано з попереднім пунктом (сильна кореляція між Recall і адекватністю для зважених N-грам): використання 1 ручного перекладу із однаковими результатами означає, що немає більше “проблем із Recall” [1] – здатність системи уникати появи надмірної кількості N-грам може зараз надійно вимірюватися. Таким чином, стало також можливо обчислювати F-бали. В результаті обчислення стали набагато простіше: Recall було знайдено, щоб дати хорошу оцінку для адекватності, а F-бали надійно прогнозують логічність.

Звичайно, використання єдиного ручного довідкового перекладу замість багатьох перекладів розширить можливості використання оцінки на основі N-грам.

Крім того, це пропонує нову лінгвістичну інтерпретацію природи цих двох якісних критеріїв: це інтуїтивно зрозуміло, що логічність відносить до категорії, тобто



передбачає адекватність (так само як F-бал відносить до категорії Recall). F-бали сильніше корелюються з адекватністю, ніж обидва з їх компонентів: точність і Recall; так само адекватність може робити внесок в логічність разом з деякими іншими коефіцієнтами. зрозуміло, що людині потрібні адекватні переклади (або як мінімум переклади, які мають сенс) для того, щоб оцінити природність, або логічність.

Здатність знаходити сенс тексту могла би бути головною основою для оцінювання адекватності: зовсім неправильні переклади в МП - це відносно рідкісне явище. Це може бути результатом принципу - наприклад для структури тексту “другий закон термодинаміки”, – набагато важче створити деякий альтернативний смисл, ніж знищити існуюче значення в перекладі, так що більшість неадекватних перекладів є тільки нісенітницею. Проте, логічно неправильні переклади трапляються навіть рідше, ніж нелогічні, згідно тому ж принципу. Справжня різниця у балах робиться на основі сегментів, які мають сенс і можуть не бути логічними, і речей, які не мають його і про які важко сказати, чи вони логічні.

Ця пропозиція може емпірично перевірятися: якщо адекватність - це необхідна вхідна умова для логічності, то повинна бути більша розбіжність у балах логічності в текстах або частинах, які мають більш низькі бали адекватності. Емпірична оцінка цієї гіпотези - це тема майбутнього дослідження.

Слід відзначити, що представлені бали кореляції самі високі, якщо одиниця оцінки - це повна сукупність перекладів, генерованих системою МП. Для оцінки тексту, кореляція набагато нижче. Це може бути завдяки тому, що людські судження не завжди послідовні і коли людина оцінює текст, який містить фрагменти різної якості, робить це більш менш випадково. Проте, ця довільність має рівні, якщо одиниця оцінки зростає в розмірі – від тексту до сукупності текстів. Це спостереження означає, що для того, щоб отримати надійні бали на рівні тексту потрібно дозволити оцінювати на рівень нижче, ніж текст, тобто на рівні окремих сегментів, які в цілому можуть фільтрувати довільність людських інтуїтивних суджень.

Автоматичні методи оцінки, як наприклад BLEU [1], RED [2,3], метод, заснований на продуктивності аналізатора виходу МП [9], або модель зважених N-грам, запропонована тут, може бути більш послідовною в оцінюванні якості в порівнянні з людськими оцінками.

В розглянутих вище матеріалах було сказано, що модель зважених N-грам полегшує використання методів оцінки МП, так як потрібний лише один ручний (окрім одномовної сукупності текстів, які отримані лише раз і що є значно дешевшим, ніж отримати багато перекладів для кожного оцінюваного тексту).

Модель також обчислює повторну появу слів і F- бали, основані на повторній появі, що, як виявилось, пов'язано з різними аспектами якості МП. Неможливість обчислення повторної появи слова методом BLEU була причиною “однобокої оцінки” і це створювало зайві ускладнення. Розділення процедур вимірювання точності і повторної появи заходів може також мати яскраву лінгвістичну інтерпретацію, і сприяти більш глибокому розумінню природи і зв'язків між різними аспектами якості МП.

Основним для обох розглянутих проблем є стабільність балів при різних способах оцінки, коли використовуються альтернативні ручні переклади. Далі порівнюється стабільність отриманих результатів із стабільністю результатів, отриманих на основі базової моделі N-грам для одного ручного перекладу.

Для того, щоб виконати це порівняння було реалізовано спрощену BLEU систему, яка здійснювала базові підрахунки N-грам без ваги, і використовувала єдиний ручний переклад [1]. Ця система працює так само як і модель зважених N-грам, але замість використання ваги S-балів просто рахує N-грами. Це порівняння показує, що стабільність оцінювання балів покращено за рахунок використання ваги, яка відображає значимість слова в тексті .

На цьому етапі експерименту вимірювались зміни, які відбуваються із балами систем МП, при використанні альтернативного ручного перекладу, як для підрахунків базових N-грам так і для моделі зважених N-грам. Середнє квадратичне відхилення було обчислено для кожної пари оціночних балів, вироблених двома рухами системи з альтернативними ручними перекладами. Середня величина цих середніх квадратичних відхилень - це міра стабільності для даного балу.

Результати цих обчислень представлені в табл. 1.4. Вверху кожної комірки “Sc” (бал) є результатом для “ ручного перекладу” , вверху комірки – результат “Перекладу експертом”. Колонки для Середніх Квадратичних Відхилень групуються. Рядок “Improved” представляє відсоток зміни в стабільності для моделі зважених N-грам.

Таблиця 1.4 - Стабільність балів зважених N-грам

| P        | Системи   | Sc-Baseline | StDev-basln | StDev-wtd | Sc-Weightd    |
|----------|-----------|-------------|-------------|-----------|---------------|
|          | candid    | 0.6364      | 0.0034      | 0.0017    | 0.6996        |
|          |           | 0.6412      |             |           | 0.7020        |
|          | globalink | 0.5449      | 0.0014      | 0.0008    | 0.6306        |
|          |           | 0.5469      |             |           | 0.6318        |
|          | ms        | 0.5295      | 0.0006      | 0.0011    | 0.6217        |
|          |           | 0.5287      |             |           | 0.6201        |
|          | reverse   | 0.6030      | 0.0018      | 0.0008    | 0.6793        |
|          |           | 0.6056      |             |           | 0.6805        |
| systran  | 0.6072    | 0.0025      | 0.0013      | 0.6850    |               |
|          | 0.6107    |             |             | 0.6869    |               |
| Ave SDev |           |             | 0.0019      | 0.0012    |               |
| Improved |           |             |             |           | <b>+36.8%</b> |

Продовження таблиці 1.4

|                     | Системи   | Sc-Baseline | StDev-basln | StDev-wtd | Sc-Weightd    |
|---------------------|-----------|-------------|-------------|-----------|---------------|
| R                   | candid    | 0.6015      | 0.0011      | 0.0004    | 0.5067        |
|                     |           | 0.6000      |             |           | 0.5072        |
|                     | globalink | 0.5564      | 0.0025      | 0.0005    | 0.4883        |
|                     |           | 0.5529      |             |           | 0.4876        |
|                     | ms        | 0.5929      | 0.0048      | 0.0029    | 0.5152        |
|                     |           | 0.5861      |             |           | 0.5111        |
|                     | reverse   | 0.6192      | 0.0025      | 0.0008    | 0.5400        |
| 0.6157              |           | 0.5389      |             |           |               |
| systran             | 0.6285    | 0.0018      | 0.0003      | 0.5511    |               |
|                     | 0.6259    |             |             | 0.5507    |               |
| Ave SDev            |           |             | 0.0025      | 0.0010    |               |
| Improved            |           |             |             |           | <b>+60.0%</b> |
| F                   | candid    | 0.6184      | 0.0011      | 0.0008    | 0.5877        |
|                     |           | 0.6199      |             |           | 0.5889        |
|                     | globalink | 0.5506      | 0.0005      | 0         | 0.5504        |
|                     |           | 0.5499      |             |           | 0.5504        |
|                     | ms        | 0.5594      | 0.0025      | 0.0023    | 0.5635        |
|                     |           | 0.5559      |             |           | 0.5603        |
|                     | reverse   | 0.6110      | 0.0003      | 0.0001    | 0.6017        |
|                     |           | 0.6106      |             |           | 0.6015        |
| systran             | 0.6177    | 0.0004      | 0.0004      | 0.6108    |               |
|                     | 0.6182    |             |             | 0.6113    |               |
| Ave SDev            |           |             | 0.0009      | 0.0007    |               |
| Improved            |           |             |             |           | <b>+22.2%</b> |
| All scores improved |           |             |             |           | <b>+39.7%</b> |

Базовий підхід дає відносно стабільні результати: середнє квадратичне відхилення не перевищувало 0.005, що означає, що BLEU дає надійні розрахунки тільки при використанні єдиного ручного перекладу (хоча інтерпретація балів для єдиного та багатьох ручних перекладів повинна бути різною).

Однак, модель зважених N-грам покращила стабільність базової моделі N-грам, навіть більше того: середнє квадратичне відхилення не перевищило 0.003, і бали в середньому на 40% більш стабільні.

### 1.5 Аналіз недоліків існуючих методів оцінювання якості машинного перекладу

Проведений аналіз сучасних методів оцінювання систем машинного перекладу, показав, що існує досить багато автоматичних методів для оцінювання різних аспектів якості МП таких, як адекватність, логічність і інформативність, які забезпечують альтернативи дорогому і тривалому процесу оцінки МП людиною. Вони здатні давати оцінки, які корелюють із оцінками якості перекладу людини-експерта. Кілька таких автоматичних методів було запропоновано в останні роки. Зокрема метод Papineni - BLEU, який заснований на порівнянні кількості збігів N-грам.

Проте серйозною проблемою для методу BLEU є велика кількість варіантів перекладів людиною. BLEU схильний вважати помилкою будь-яке відхилення від базового набору N-грам, не дивлячись на те, що звичайно існує декілька прийнятних варіантів перекладу однієї і тієї ж частини тексту, і тому будь-які відхилення в якості МП можуть бути тільки частково віднесені до помилок. Звичайно, потрібно багато різних варіантів перекладу експертів для одного і того ж тексту для того, щоб вирішити цю проблему, але це робить автоматичну оцінку більш дорогою.

Результати для зважених моделей N-грам мають істотно вищу кореляцію із оцінками якості перекладу людини-експерта про адекватність, логічність і інформативність, ніж базові підходи оцінки N-грам, які використовуються методом BLEU. Метод оцінки на N-грамах понині береться за стандарт, проте він використовує лише статистичні дані, і не враховує синтаксичну інформацію про мову.

Було проведено оцінку якості МП на основі лексичного, граматичного і синтаксичного розбору і статистичного аналізу помилок в текстах перекладених системою PROMT, яка є найкращою з відомих систем.

1. При роботі з даним програмним продуктом були виявлені наступні погрішності при перекладі тексту. Лексичний аналіз тексту показав, що PROMT в більшій частині правильно переводить прості частини мови, але допускає помилки в перекладі відмінків, приналежності прикметників, мовних оборотів, побудови речень. Неточно перекладаються слова, що мають кілька значень. Граматичний аналіз тексту показав, що електронний перекладач справляється з перекладом слів в множині й однині, але є певні труднощі в перекладі відмінків і постановці дієслів у потрібне число. Це пояснюється різною інтерпретацією відмінків у російській і англійській мовах. Статистичний аналіз помилок перекладеного тексту, здійсненого в англо-російському і російсько-англійському напрямках, було виконаного для його репрезентативної вибірки обсягом в 250 слів [38]. Він показав, що при перекладі в російсько-англійському напрямку синтаксичні помилки виявлено в 80% речень, граматичних помилок – 28,6%, лексичних помилок – 4,5%. При перекладі в англійсько-російському напрямку виявлено 18% синтаксичних, 86% граматичних і 35,6% лексичних помилок.

## 1.6 Висновки

Таким чином, на сьогоднішній день існує велика кількість моделей та підходів, щодо оцінювання МП. Кожна з них має свої переваги та недоліки, і її ефективність залежить від граматики та семантики мови. При цьому кожна з таких моделей концентрує увагу на статистичних показниках мови, а лінгвістична інформація залишається поза увагою.

Серйозна проблема для методу BLEU - це необхідність великої кількості варіантів перекладів людиною. BLEU схильний вважати помилкою будь-яке відхилення від базового набору N-грам, не дивлячись на те, що звичайно існує декілька прийнятних варіантів перекладу одного тексту, і тому будь-які відхилення в якості МП можуть бути тільки частково віднесені до помилок.

Метод оцінки якості перекладу на основі зваженої моделі N-грам, дещо виправляє помилки методу BLEU, акцентуючи увагу на тому, що деякі фрагменти інформації в тексті мають більш важливе - ключове значення і тому мають зберігатись в будь-якому перекладі, менш важлива інформація і певні функціональні слова підлягають більшій варіації, оскільки перекладач має більшу свободу вибору.

Зважаючи на результати аналізу існуючих методів оцінки МП, залишається актуальною проблема створення вдосконалених методів автоматичного оцінювання систем машинного перекладу, які краще враховують лінгвістичну інформацію та знаходяться на одному рівні із людським оцінюванням.

## 2 РОЗРОБКА ВДОСКОНАЛЕНИХ МЕТОДІВ ОЦІНЮВАННЯ СИСТЕМ МАШИННОГО ПЕРЕКЛАДУ МОВИ

Проведений аналіз показав, що оцінювання якості машинного перекладу за допомогою людини є якісним, але достатньо затратним. Існуючі методи автоматичного оцінювання неповною мірою враховують лінгвістичну інформацію (семантику, граматику). Більшість методів використовують процедуру послівного оцінювання. Така процедура дає прийнятні за точністю результати тільки для текстів нефлективних мов. Задачею даного розділу є розробка нової моделі процесу перекладу та створення методу автоматичного оцінювання якості систем МП, який враховує вказані недоліки.

### 2.1. Модифікований метод N-грам із зміненою точністю та використанням морфем

Основним критерієм для обчислення якості систем машинного перекладу мови є міра точності. Щоб обчислювати точність, зазвичай просто підраховується число слів в МП (unigrams), які містяться в будь-якому перекладі ПЕ, і потім ділиться на загальну кількість слів у машинному перекладі.

За цим методом точність перекладу підраховується за формулою [1]:

$$P_n = \frac{\sum_{C \in (\text{Candidates})} \sum_{n\text{-gram} \in C} \text{Count}_{clip}(n\text{-gram})}{\sum_{C' \in (\text{Candidates})} \sum_{n\text{-gram}' \in C'} \text{Count}_{clip}(n\text{-gram}')}, \quad (2.1)$$

де в чисельнику підсумовується кількість N-грам машинного перекладу, що співпали з N-грамами еталонного перекладу, а в знаменнику обчислюється загальна кількість N-грам еталонного перекладу.

На жаль, система оцінки може кілька разів порівнювати слова які повторюються, показуючи недостовірну високу точність перекладу, як це показано у прикладі нижче. Інтуїтивно проблема проста: слово порівняння не потрібно розглядати ще раз, якщо до цього відповідне слово МП ідентифіковане. Врахувавши це, отримаємо модифікований метод N-грам із зміненою точністю. Щоб врахувати це, спочатку рахуємо скільки разів слово повторюється в будь-якому перекладі експерта – це число буде максимальною

кількістю співпадінь даного слова, потім ділимо на загальну кількість слів у МП. Пояснимо це таким прикладом.

Приклад 2:

МП: the the the the the tha the

ЕП 1: The cat is on the mat.

ЕП 2: There is a cat on the mat.

Отже модифікована точність Unigram =  $2/7$ .

У Прикладі 1 (глава 1.4.1), МП 1 досягає модифікованої unigram точності  $17/18$ ; тоді як МП 2 досягає модифікованої unigram точності  $8/14$ . Точно так само модифікована unigram точність у Прикладі 2 -  $2/7$ , навіть при тому, що його стандартна unigram точність -  $7/7$ .

Модифікована точність N-грама обчислюється аналогічно для будь-якого N: вся кількість N-грама МП і їх відповідна максимальна кількість в ПЕ підраховані. Кількість співпадінь N-грам підраховується, і ділиться на загальну кількість N-грам МП. У Прикладі 1, МП 1 досягає зміненої точності 2-грам  $10/17$ , тоді як менш якісний МП 2 досягає зміненої bigram точності  $1/13$ . У Прикладі 2, МП досягає зміненої bigram точності 0. Метод модифікованої точності враховує два аспекти перекладу: адекватність і швидкість. Переклад, використовуючи ті ж самі слова (1 грам) як у ЕП задовольняє адекватність. Більш довгий N-грам, ще й до того ж дає змогу оцінити смислові характеристики МП

Метод оцінки на N-грамах понині береться за стандарт, проте він використовує лише статистичні дані послівно, і не враховує синтаксичну інформацію про мову.

Практика оцінювання флективних мов [25] (в яких у вираженні граматичних значень провідну роль відіграє флексія) показує, що використання тільки послівного аналізу для їх оцінювання не дозволяє отримати задовільних результатів, тому виникає необхідність використання додаткової інформації. Природно за таку інформацію використати контекстну, в якості якої може слугувати лексична і статистична інформація. Лексичну інформацію зручно використовувати тоді, коли елементами розпізнавання будуть морфи – мінімальні змістовно розділимі частини слова (наприклад, префікс, корінь, суфікс, закінчення). Тоді можна використати статистичну і лексичну інформацію про морфи для побудови моделей слів тексту у



вигляді прихованих марковських мереж (ПММ), що дозволило б застосувати відомі алгоритми на цих мережах [67].

Як випливає з наведених міркувань, актуальною є задача розробки програмного забезпечення для побудови бази даних українських морфем для подальшого використання в задачі оцінювання МП. З набору текстів, набраних з різних джерел української мови (художні тексти, науково-публіцистичні, технічні тощо) і достатніх за розміром для утворення навчальної вибірки, необхідно виділити повну множину морфем українського мови.

Розробка суттєво нового програмного забезпечення для побудови бази даних морфем української мови вимагає великих часових затрат та зусиль великого колективу. Тому для розв'язання поставленої задачі застосуємо вільно розповсюджуваний програмний продукт „Morfessor1.0.pl” [68] до текстів на російській та українській мовах. Отримана в результаті доопрацювання програма „Морфесор 1.0” [69] дозволяє повністю автоматизувати роботу згаданої вище програми для проведення морфологічного аналізу (розбиття слів на елементарні частини – префікси, суфікси, корені і т.д.). Також дана програма дозволяє перевірити адекватність утвореної лінгвістичної моделі тексту.

В математичному плані побудова лінгвістичної моделі з заданої для навчання вибірки тексту (корпуса) базується на принципі максимуму апостеріорної оцінки повної імовірності. Ця модель складається з словника морфем (лексикона) і граматики. Оптимальна модель повинна під час сегментації текстового корпусу на морфеми давати мінімальний із можливих набір морфем. Максимум апостеріорної оцінки імовірності параметрів лінгвістичної моделі знаходиться, як:

$$\arg \max_M P(M / corp) = \arg \max_M P(corp / M) \cdot P(M), \quad (2.2)$$

де  $P(M)$  - імовірність лінгвістичної моделі,  $P(corp / M)$  - імовірність корпусу тексту, який залежить від даної лінгвістичної моделі.

Імовірність лінгвістичної моделі  $P(M)$  - це сумісна імовірність введеного словника і граматики:

$$P(M) = P(\text{lexicon}, \text{grammar}). \quad (2.3)$$

Лексикон містить по одному з екземплярів кожної морфеми з просегментованого

корпуса тексту. Властивості, які зберігаються в лексиконі, - це частота (кількість спостережень) морфеми в корпусі, а також рядок букв, з яких складається морфема. Пошук морфем лексикона здійснюється шляхом знаходження статистично стійких N-грам в корпусі тексту [68].

Грамматика містить інформацію про те, як можуть комбінуватися блоки лексикону. Відсутність граматики говорить про те, що морфема не залежить від того, яка морфема передує їй чи йде за нею, і чи вона знаходиться на початку, посередині чи в кінці слова.

На відміну від вихідної програми, яка потребує попередньої рутинної підготовки даних у вигляді однослівних стрічок, „Морфесор 1.0” проводить автоматизовану підготовку довільного тексту до вигляду, придатного для морфологічного аналізу. Дана програма без втручання користувача:

- вилучає зі вхідного тексту усі символи, що не є літерами;
- записує кожне окреме слово у новий рядок текстового файлу.

Приклад роботи процедури підготовки вхідного тексту до морфологічного аналізу:

а) вхідний текст

Я народився у 1927 році - єдиний син небагатих англійців, яким до самої смерті не удавалося вирватися за межі тіні виродливої карлиці, королеви Вікторії.

б) відформатований текст

Я  
народився  
у  
році  
єдиний  
син  
небагатих  
англійців  
яким  
до  
самої  
смерті  
не  
удавалося  
... ..

Графічний інтерфейс програми “Морфесор 1.0” показаний на рис. 2.1. У головному вікні програми користувач має вказати ім’я текстового файлу, який необхідно сегментувати на морфеми. Також користувач повинен вказати директорію для збереження та ім’я результуючого файлу. Програма автоматично призначає ту саму директорію для результуючого файлу, де знаходиться вхідний файл. Також програма генерує ім’я результуючого файлу, шляхом додавання до імені вхідного файлу службового слова «Results» та поточної дати.

Як тільки модель морфемної сегментації навчена на деякому наборі даних, вона може використовуватися для сегментування нових словоформ. У цьому режимі сегментації програми “Морфесор” не відбувається навчання моделі. Кожне слово на вході сегментується на морфеми за допомогою алгоритму Вітерці [68], що знаходить самі схожі сегменти слова в послідовності морфем, що є в існуючій моделі. (для того, щоб забезпечити можливість існування, як мінімум, однієї сегментації, кожен окремий символ слова, що не існує, як морфема, може розглядатися як морфема з дуже малою імовірністю). Користувач може ввести ім’я файлу, що містить існуючу модель (друге вікно графічного інтерфейсу на рис. 2.1.).

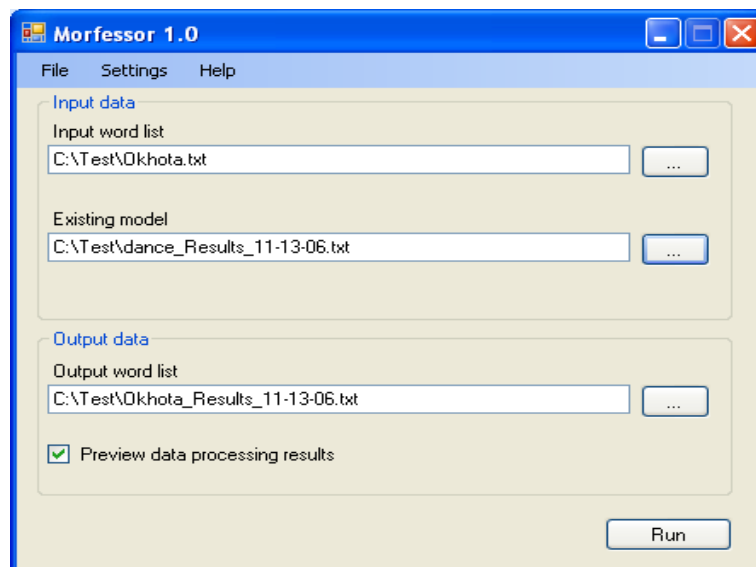


Рисунок 2.1 - Графічний інтерфейс програми “Морфесор 1.0”

“Морфесор” записує вихідні дані у текстовий файл, введений користувачем. Деякі рядки виходу програми починаються із символу #. Ці рядки можуть розглядатися як коментарі і включати, наприклад, всю службову інформацію

(статистика морфем, час виконання та ін.). Головна частина виходу складається з морфемних сегментацій слів вхідного тексту [66]. Використовуваний формат впливає із наступних морфемних сегментацій деяких слів. Попереду кожного слова вказується частота появи в тексті:

...

1 Ю + ний

6 Александр

2 за + во + ював

1 Ін + ді + ю

15 Не + вже

330 сам

1 Це + за + р

4 роз + бив

1 га + л + лі + в

546 чи

368 мав

257 коло

326 себе

117 бодай

1 кухар + я

1 Ф + і + ліп

1 і + с + пан + ський

18 плакав

...

Перевірка адекватності утвореної лінгвістичної моделі здійснюється наступним чином.

Після завершення виконання морфологічного аналізу програма пропонує користувачеві протестувати утворену лінгвістичну модель та перевірити її адекватність.

Адекватність моделі визначалась наступним чином. Визначалась вибірка довільного тексту (660 слів), далі текст було сегментовано на морфеми “вручну” та з

сегментованого тексту було вилучено усі морфеми, що повторюються два та більше разів. В результаті було отримано 522 морфеми. Ці морфеми було взято за еталон. Після виконання сегментації вхідного тексту, програма «Морфесор 1.0» проганяє ту саму сторінку тестового тексту (660 слів) з використанням на вході у якості існуючої моделі щойно утвореної лінгвістичної моделі. Далі із сегментованого тестового тексту вилучаються морфеми, що повторюються. Для перевірки адекватності взято критерій невідповідності морфем тестового тексту та еталонного: проганяється масив морфем тестового тексту по кожній морфемі, якщо дана морфема присутня в еталонному наборі, то помилка не враховується, в протилежному випадку – помилка зараховується.

Аналіз практичних результатів оцінювання систем МП показав, що використання послівного аналізу текстів перекладених машиною не дозволяє отримати задовільних результатів, тому виникає необхідність інтелектуалізації методів оцінювання МП. Для цього в роботі запропоновано використовувати поморфемне оцінювання тексту, яке дозволяє використати змістовну (лексичну і лінгвістичну), та модифіковано програмне забезпечення для побудови бази морфем української мови.

## 2.2. Метод оцінки ефективності систем МП на основі інформаційно-статистичного критерію і моделі потенціальної системи

Метод оцінки на N-грамах в більшій мірі підходить для систем статистичного МП, оскільки не враховує синтаксичні і семантичні аспекти мови. Автори в даному пункті описують шлях реалізації запропонованого ними в [38] методу оцінки ефективності систем МП, що ґрунтується на інформаційно-статистичному критерії і моделі потенціальної системи [70]. Цей критерій, поряд із статистичними характеристиками мови, враховує і її синтаксичну структуру.

Грамматика представляє собою поверхневу структуру мовного повідомлення і містить фонетичну структуру слів та порядок слів в реченні [71,72].

В вибраному підході одним із компонентів методу оцінки якості систем машинного перекладу є набір речень на вихідній мові, отриманих високо-професіональними перекладачами в результаті перекладу тексту вхідної мови. Будемо вважати їх реченнями мови  $L(G)$ , яка генерується граматикою  $G$ . Процес перекладу

розглядається як процес передачі повідомлення, зашифрованого текстом вхідної мови, по каналу з шумами [73]. Шуми каналу відповідають огріхам в перекладі тексту вихідною мовою. В такому випадку якість машинного перекладу автори пропонують оцінювати за допомогою модифікованого до систем штучного інтелекту інформаційного критерію [40]:

$$E = \frac{I_{MT}}{I_p}, \quad (2.4)$$

де  $I_{MT}$  - інформативність перекладеного машинним перекладачем тексту,  $I_p$  - інформативність вхідного тексту. Вони можуть бути визначені з таких виразів:

$$I_{MT} = H(L(G)) - H(L(G))_{MT}, \quad (2.5)$$

$$I_p = H(L(G)) - H(L(G))_p, \quad (2.6)$$

де  $H(L(G))$  - ентропія мови;  $H(LG))_{MT}$  - ентропія тексту, перекладеного машинним перекладачем;  $H(L(G))_p$  - ентропія тексту, перекладеного потенціальною (ідеальною) системою перекладу. В якості потенціальної системи перекладу вибрано людину – професіонального перекладача, тому в першому наближенні можна вважати, що  $H(L(G))_p = H(L(G))$ .

Такий підхід, на відміну від запропонованого в роботі [1], дозволяє оцінити основний параметр якості перекладу – його інформативність. Для реалізації цього підходу необхідно визначити шляхи обчислення величин ентропії  $H(L(G))$  і  $H(LG))_{MT}$ . З цією метою будемо розглядати мову  $L(G)$  у вигляді діаграми переходів між станами без наявності в них петель. Діаграма переходів містить скінчений набір станів  $Q$ , скінчений словник  $V$  і функцію переходів  $\delta$ , яка встановлює граматичні обмеження:

$$\delta: (Q \times V) \rightarrow Q, \quad (2.7)$$

тобто кожен перехід на діаграмі має вигляд:

$$\delta(q_i, v) = q_j. \quad (2.8)$$

При цьому стани  $q_i \in Q$  вказують на порядок слів, а  $v(q_{ij})$  - самі слова. Правильне речення  $w$  складається з міток всіх дуг, що проходять послідовно вздовж будь-якого шляху з початком в  $q_1$  і кінцевим станом  $q_l \in Z$ , де  $Z$  - множина кінцевих станів.

Визначальною кількісною величиною, яка характеризує діаграму, є її матриця зв'язності  $C$ , кожен елемент якої  $c_{ij}$  дорівнює числу переходів (дуг)  $r$  від стану  $q_i$  до стану  $q_j$ :

$$c_{ij} = r, \quad i, j = 1, \vec{N}, \quad (2.9)$$

де  $r$  - число переходів з стану  $q_i$  в стан  $q_j$ .

Матриця зв'язності використовується для обчислення кількості речень  $N_k$  довжиною  $k$  в мові  $L(G)$  з виразу:

$$N_k = \vec{e}_1 C^k \vec{f}^T, \quad (2.10)$$

де  $\vec{e}_1$  представляє перші  $N$  компонент одиничного вектора;

$\vec{f}$  -  $N$ - компонентний вектор одиниць або нулів,  $i$ -та компонента якого дорівнює 1, якщо  $q_i$  є кінцевим станом.

Оскільки

$$\sum_{k=0}^{\infty} C^k = (I - C)^{-1}, \quad (2.11)$$

де  $I$  - одинична матриця розміром  $N \times N$ , то загальна кількість речень в мові  $|L(G)|$  дорівнює

$$|L(G)| = \vec{e}_1 (I - C)^{-1} \vec{f}^T \quad (2.12)$$

і середня довжина речення:

$$|\bar{W}| = \frac{\sum_k k N_k}{|L(G)|}. \quad (2.13)$$

При умові  $|L(G)| < \infty$  і рівноймовірності речень можна розрахувати  $N_k$ ,  $|L(G)|$ ,  $|\bar{W}|$  і ентропію мови:

$$H(L(G)) = \frac{\log_2 |L(G)|}{|\bar{W}|}. \quad (2.14)$$

Якщо умова рівноймовірності речень не виконується, то можна отримати максимальну ентропію при будь-якому імовірнісному розподіленні речень:

$$H_{\max}(L(G)) = -\log_2(x_0), \quad (2.15)$$

де  $x_0$  обрано з того розрахунку, щоб  $1 - \sum_k N_k x_0^k = 0$ .

Визначивши ентропію мови, можна визначити і ентропію тексту  $H(LG)_{MT}$ , отриманого машинним перекладачем. Для цього система автоматичного перекладу мови представляється у вигляді каналу передачі інформації з шумами [73-78] (рис. 2.2.).

Основною характеристикою даного каналу є втрата інформації на одне повідомлення (речення)  $H(w/w^*)$ , яка є мірою невизначеності перекладу речення  $w$  в речення  $w^*$ . Звідси ентропію перекладу  $H(LG)_{MT}$ , отриманого машинним перекладачем, можна отримати з виразу:

$$H(L(G))_{MT} = H(L(G) | H(w/w^*)), \quad (2.16)$$

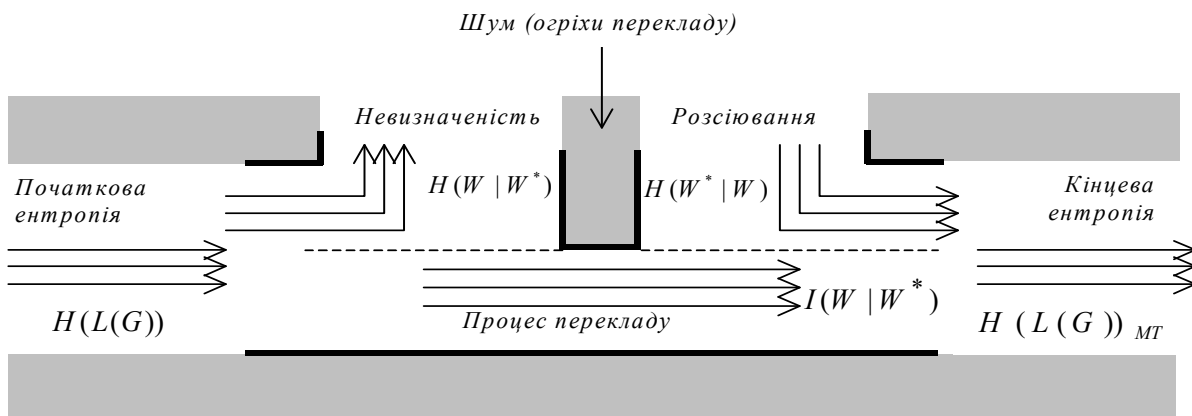


Рисунок 2.2 - Представлення процесу машинного перекладу тексту каналом з шумами

тобто втрата інформації в перекладеному тексті дорівнює добутку ентропії на одне речення на середню кількість речень в тексті[80]:

$$H(w/w^*) = -\sum P(w_i/w_i^*) \cdot \log P(w_i/w_i^*). \quad (2.17)$$



Величини  $p(w_i / w_i^*)$ , що характеризують подібність речень, можуть бути визначені з формули [2.14]:

$$p(w_i / w_i^*) = \frac{\tilde{\mu}_i \cdot 10^{-d_{ii}^w}}{\sum_{k=1}^{|\tilde{w}|} \tilde{\mu}_k \cdot 10^{-d_{ik}^w}}, \quad (2.18)$$

де  $d_{ii}^w$  - модифікована відстань Левенштейна між  $i$ -им еталонним реченням і  $i$ -тим перекладеним реченнями тексту, яка визначається як сумарна кількість пропусків, замінів, вставок та змін порядку слів, яку потрібно зробити у перекладеному реченні, щоб отримати еталонне, а  $\tilde{\mu}_k$  - середня величина відстані між реченнями [81].

Для підтвердження адекватності розробленого методу було обчислено значення критерію (2.4) для перекладу 120 еталонних речень, виконаного двома перекладачами-людьми Н1 і Н2 та двома системами машинного перекладу МТ1 і МТ2. Результати розрахунків представлені на рис. 2.2.

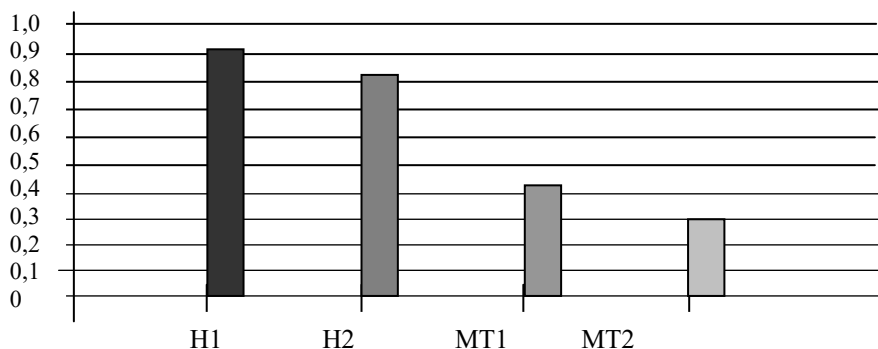


Рисунок 2.3 - Якість перекладу згідно вибраного критерію

В роботі [41] було проведено експертну оцінку цих же перекладів професіональними лінгвістами, яка показала однаковий з отриманим ранговий порядок їх суджень про якість роботи перекладачів. Таким чином, попередній аналіз запропонованого методу показав можливість його застосування для автоматичної оцінки систем машинного перекладу мови.

Враховувавши модифікований метод N-грам із зміненою точністю та використанням морфем, і метод оцінки ефективності систем МП на основі інформаційно-статистичного критерію і моделі потенціальної системи, пропонується така узагальнена схема оцінювання:

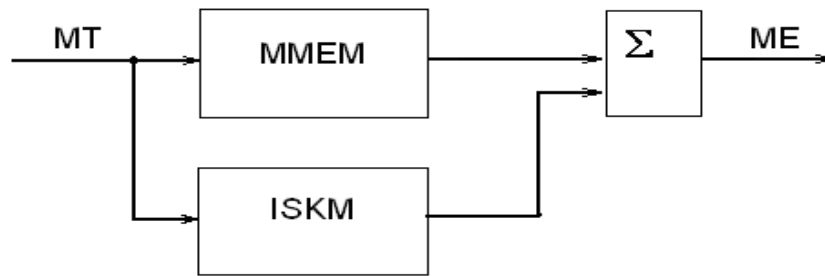


Рисунок 2.4 - Узагальнена схема оцінювання систем МП

де МТ – машинний переклад;

МЕ – автоматична оцінка машинного перекладу;

ММЕМ - модифікований метод N-грам із зміненою точністю та використанням морфем;

ISKM - метод оцінки ефективності систем МП на основі інформаційно-статистичного критерію і моделі потенціальної системи.

### 2.3. Висновки

На основі виконаних теоретичних досліджень розроблено рекомендації з удосконалення методик оцінювання систем автоматичного перекладу мови.

Удосконалено метод статистичного оцінювання систем автоматичного перекладу мови, який відрізняється від існуючих тим, що враховує повторні появи N-грам у реченні, що дає змогу більш точно, поморфемно оцінити кількість їх збігів у еталонному та машинному перекладі.

Запропоновано нову модель процесу перекладу мови, яка відрізняється тим, що розглядає його, як процес передачі повідомлення зашифрованого текстом вхідної мови по каналу з шумами, що дає змогу оцінити основний параметр якості перекладеного тексту – його інформативність.

Розроблено метод оцінки ефективності систем машинного перекладу, що ґрунтується на інформаційно-статистичному критерії і моделі потенціальної системи, який відрізняється тим, що поряд із статистичними характеристиками мови, враховує її синтаксичну структуру, що дає змогу оцінити не тільки статистичну, а й синтаксичну інформацію про мову.

### 3 РОЗРОБКА АЛГОРИТМІВ І ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Підсумковим етапом математичного моделювання є встановлення відповідності розробленої математичної моделі реальним процесам, що протікають в досліджуваних об'єктах чи системах. Якщо результати експериментальних досліджень моделі збігаються з теоретичними, то це підтверджує її адекватність об'єкту.

В даному розділі проведена розробка програмного забезпечення для виконання натурального експерименту із оцінюванням систем машинного перекладу мови на основі модифікованого методу порівняння кількості збігів N-грам, а також перевіряється адекватність розробленої моделі.

#### 3.1 Розробка алгоритмів для реалізації модифікованого методу порівняння кількості збігів N-грам

Для програмної реалізації модифікованого методу на N-грамах, потрібні початкові дані, тобто текст перекладений експертом та текст, перекладений системою машинного перекладу.

Для прикладу візьмемо фрагмент англійського тексту Wallaby:

“The wallaby is a species of kangaroo. Like the kangaroo it lives in its mother’s pouch the first few months after birth until it can stand on its own feet. This helps it to survive the hard conditions. Wallabies can live for 8 or 9 months without fresh water. They get the water they need from plants so they don’t need to be near water.”

Переклад експерта (на російську мову) (НТ) для нього:

“Валлабу является разновидностью кенгуру. Как и кенгуру он живет в сумке матери первые несколько месяцев после рождения до тех пор пока он сможет стоять на собственных ногах. Это помогает ему выжить в тяжелых условиях. Валлабу могут жить 8 или 9 месяцев без свежей воды. Воду, которая им необходима, они получают из растений, поэтому им не нужно быть возле воды.”

Візьмемо машинний (МТ) переклад за допомогою системи Pragma:

“Кенгуру - это species кенгуру. Подобно кенгуру это живет в сумке матери первые несколько месяцев после начала, пока это может настоять на собственных

ногах. Это помощь это, чтобы выдержать трудные условия. Кенгуру могут жить в течение 8 или 9 месяцев без пресной воды. Они получают воду им нужно от заводов так что им не нужно быть возле воды.”

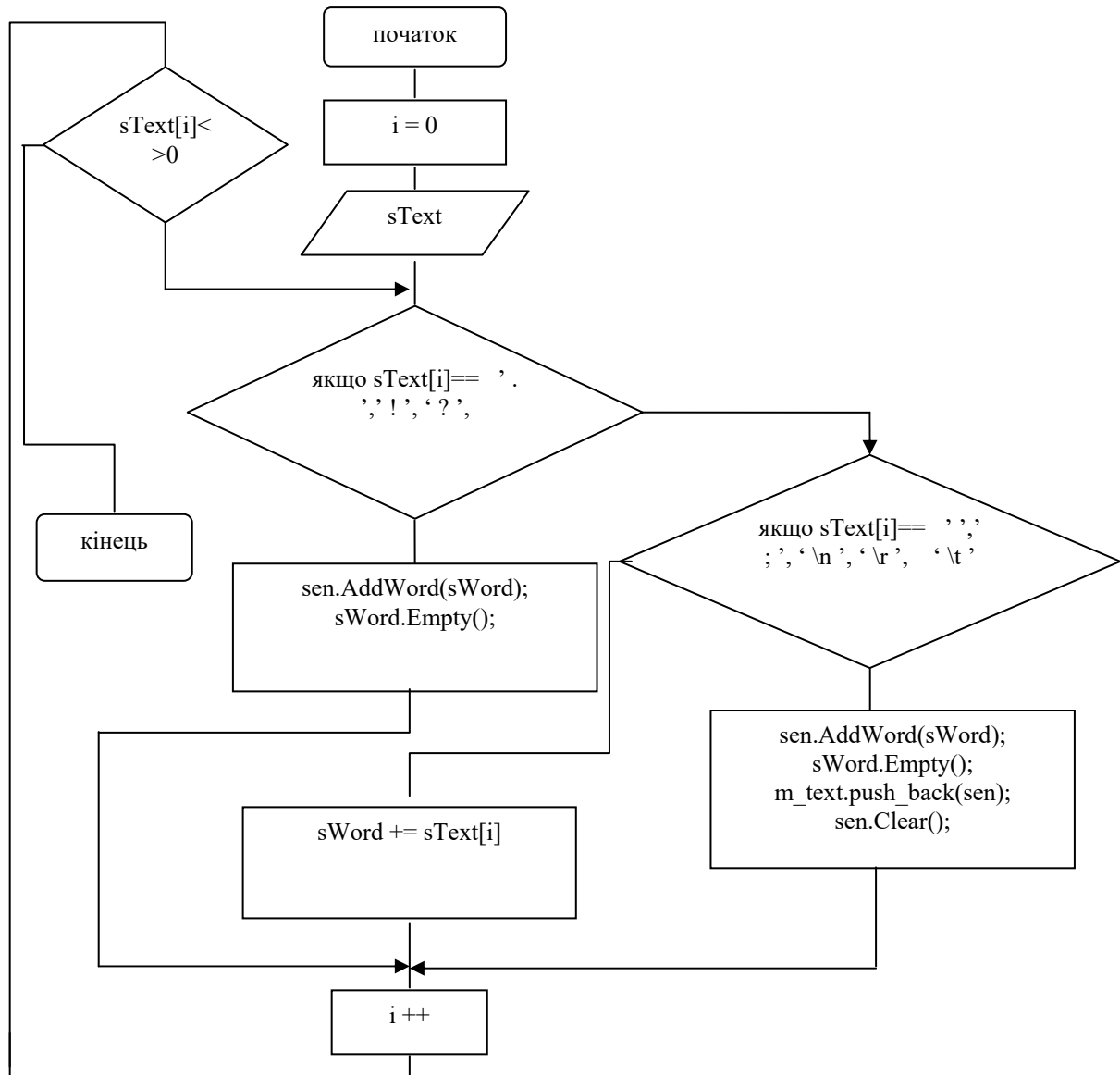


Рисунок 3.1 - Блок-схема алгоритму приведення тексту до зручного для оцінювання вигляду

Для порівняння цих перекладів машиною потрібно привести їх до вигляду зручного для аналізу, оскільки потрібно відкинути неточності у відповідності між реченнями, які виникають через розділові знаки, абзаци, надмірну кількість пропусків між словами та ін.

Для приведення до зручного для аналізу вигляду використаємо алгоритм, зображений на рис. 3.1.

Далі оцінюємо якість МП за таким алгоритмом:

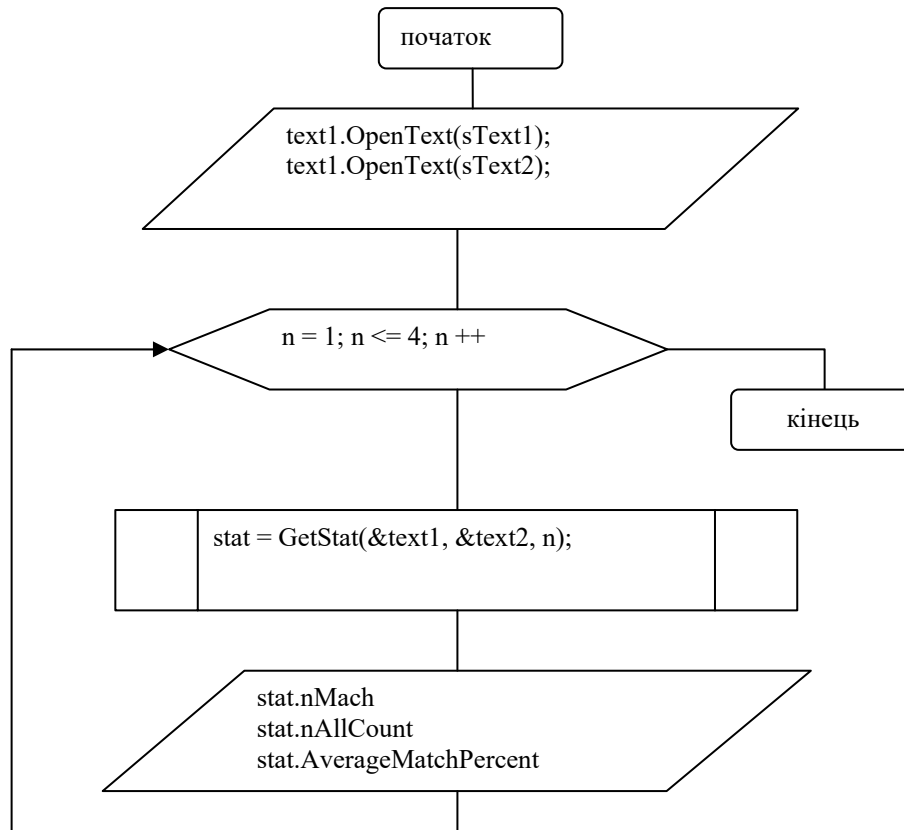


Рисунок 3.2 - Алгоритм порівняння двох текстів

Алгоритм дає змогу оцінити якість МП на основі розробленого модифікованого методу на N-грамах із зміненою точністю.

На рис 3.3. зображена структура програми. Головна програма завантажує зовнішній файл з перекладом, здійсненим машиною, і файл перекладу того ж тексту, отриманого за допомогою людини-експерта, після чого здійснює порівняння даних текстів і дає оцінку якості системи автоматичного перекладу мови.

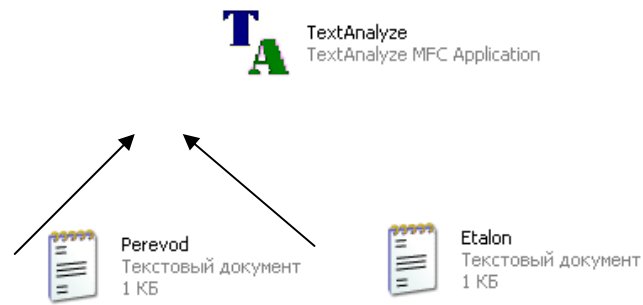


Рисунок 3.3 - Структура програми-аналізатора

На рис. 3.3. зображена структура програми-аналізатора, де:

- TextAnalyze.exe – головна програма;
- Perevod – модуль, в якому зберігається машинний переклад.
- Etalon – модуль, в якому зберігається переклад експерта.

Розглянемо файлову структуру побудови головної програми.

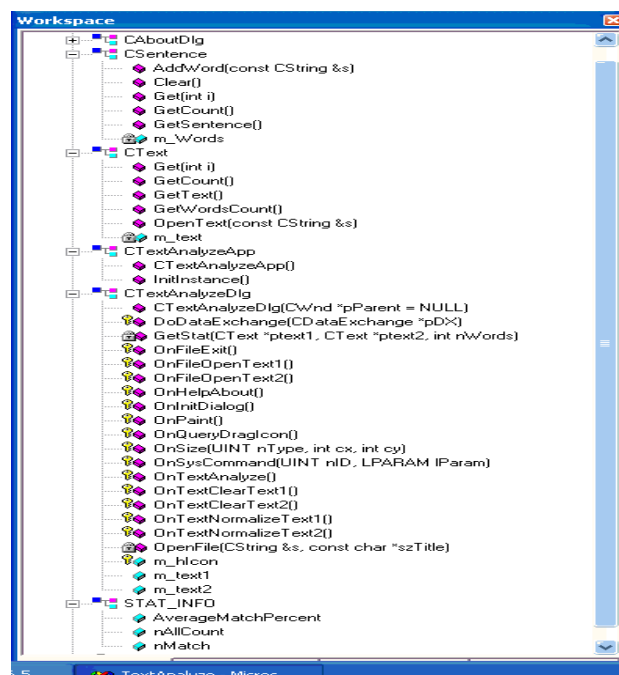


Рисунок 3.4 - Файлова структура головної програми

Опис класів та методів головної програми міститься в додатку Б.

### 3.2 Дослідження систем машинного перекладу розробленим методом

Для початку роботи програми потрібно запустити виконуючий файл TextAnalyze.exe. На екрані з'явиться вікно, представлене на рис. 3.5.

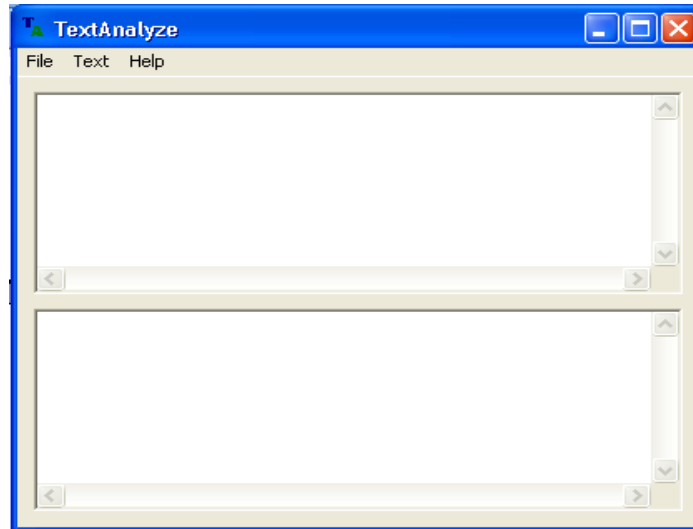


Рисунок 3.5 - Графічний інтерфейс програми

Проведемо оцінювання систем МП за допомогою розробленого програмного забезпечення на основі модифікованого методу N-грам із зміненою точністю:

Проведемо дослідження оцінювання систем МП на прикладі тексту “Wallaby”.

Для дослідження використаємо такі системи:

- Pragma;
- Prompt 7.0;
- Google Translate компанії Google;

Перекладемо текст “Wallaby” вказаними системами та оцінимо якість МП.

Результат перекладу за допомогою системи Pragma:

“Кенгуру - это species кенгуру. Подобно кенгуру это живет в сумке матери первые несколько месяцев после начала, пока это может настоять на собственных ногах. Это помощь это, чтобы выдержать трудные условия. Кенгуру могут жить в течение 8 или 9 месяцев без пресной воды. Они получают воду им нужно от заводов так что им не нужно быть возле воды.”

Оцінимо переклад програмою Pragma, для цього завантажимо Файл → OpenText1 → Pragma.txt та Файл → OpenText2 → Exp.txt, потім натиснемо Text → Analyze і отримаємо такий результат:

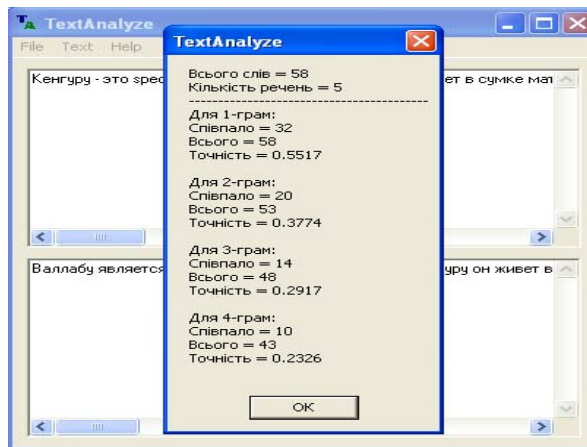


Рисунок 3.6 - Результати оцінювання системи Pragma

Отримаємо діаграму порівняння кількості збігів N-грам, натиснувши → ОК.

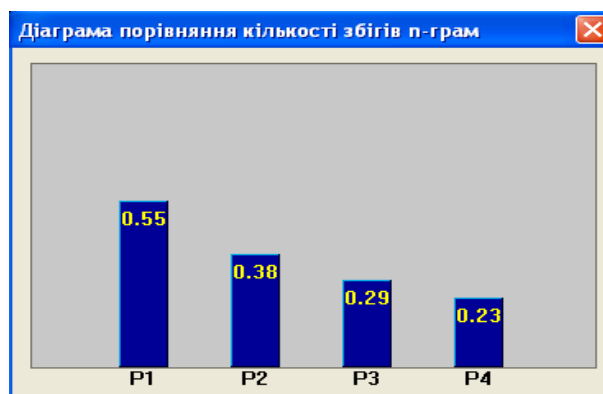


Рисунок 3.7 – Діаграма порівняння кількості збігів N-грам для системи Pragma

Результат перекладу тексту “Wallaby” за допомогою системи Promt зображено на рис. 3.8.

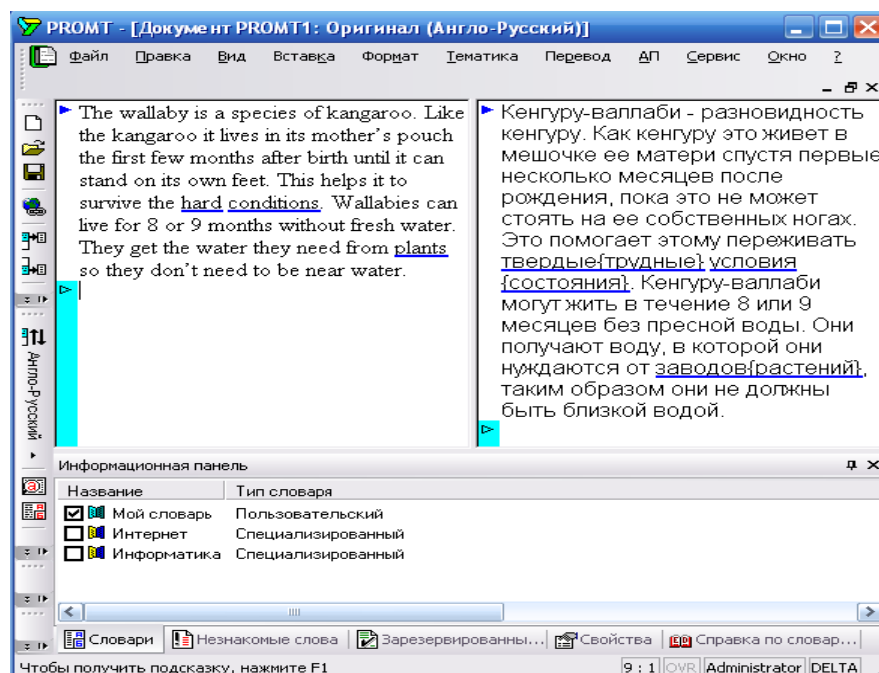


Рисунок 3.8 - Результат перекладу тексту за допомогою системи Promt



Оцінимо переклад програмою Promt, для цього завантажимо Файл → OpenText1 → Promt.txt та Файл → OpenText2 → Exp.txt, потім натиснемо Text → Analyze і отримаємо такий результат:

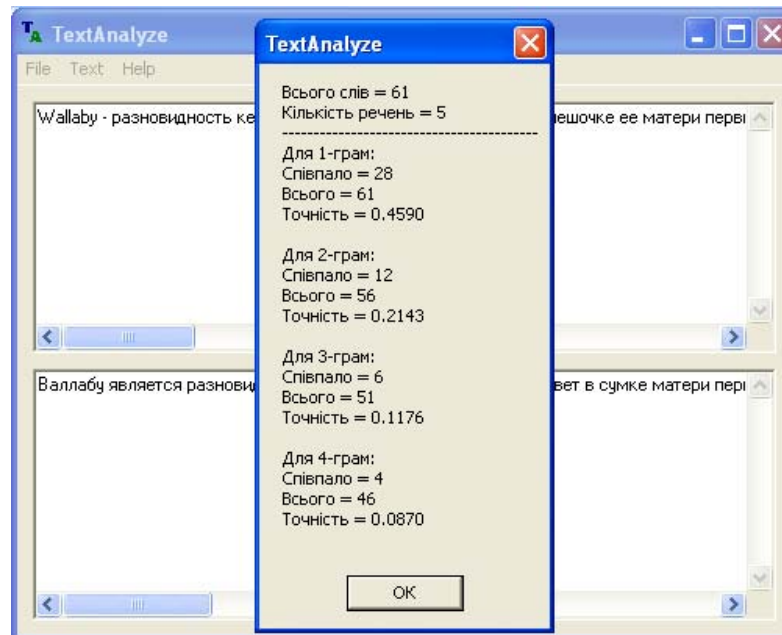


Рисунок 3.9 - Результати оцінювання системи Promt

Отримаємо діаграму порівняння кількості збігів N-грам:



Рисунок 3.10 - Діаграма порівняння кількості збігів N-грам, для системи Promt

Оцінимо переклад програмою Google Translate, для цього завантажимо Файл → OpenText1 → googl.txt та Файл → OpenText2 → Exp.txt, потім натиснемо Text → Analyze і отримаємо такий результат:

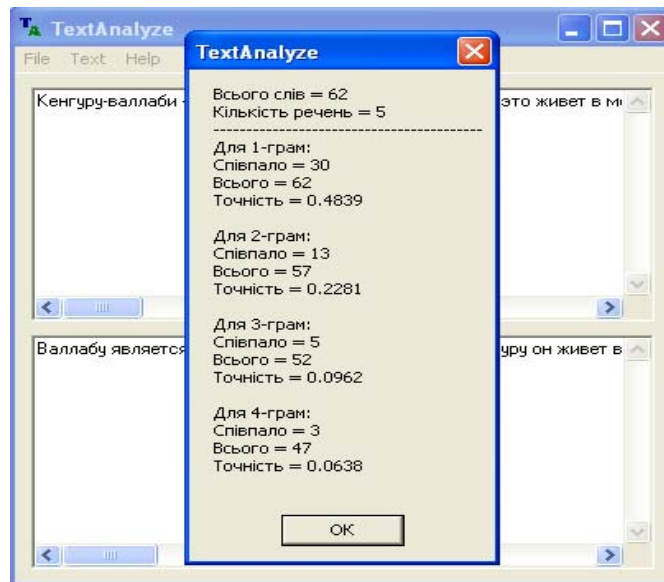


Рисунок 3.11 - Результати оцінювання системи Google Translate

Отримаємо діаграму порівняння кількості збігів N-грам, натиснувши → ОК:



Рисунок 3.12 - Діаграма порівняння кількості збігів N-грам, для системи Google Translate

За результатами оцінювання перекладу одного тексту неможливо стверджувати, що одна система є кращою за іншу, оскільки тексти бувають різноманітних тематик, тому з одними текстами може справлятися краще одна система, а з іншими - інша. Для визначення адекватності оцінювання розробленим методом протестовано 100 перекладів текстів з англійської мови, що дає змогу провести кореляційний аналіз отриманих результатів та результатів прямого оцінювання якості текстів людиною-експертом.

### 3.3 Дослідження адекватності розроблених математичних методів і моделей

Дослідження адекватності математичної моделі можна проводити двома шляхами: порівнянням результатів моделювання і результатів натурального експерименту, порівнянням результатів моделювання і результатів імітаційного (найчастіше комп'ютерного) експерименту. Бажаним, звичайно, є проведення натурального експерименту, однак його не завжди можливо провести через складність досліджуваного об'єкта, проведення натурального експерименту вимагає побудови самої системи, а відповідно і значних матеріальних витрат тому іноді доводиться обмежуватися результатами імітаційних експериментів. Що стосується експериментів по оцінюванню систем МП, то проведення натурального експерименту є можливим, оскільки системи перекладу мови розроблені і доступні для експериментування.

Алгоритм проведення натурального експерименту для дослідження адекватності математичної моделі оцінювання систем автоматичного перекладу мови складається з таких кроків:

1. Одержання результатів оцінки систем МП за допомогою розробленого методу та за допомогою експерта.
2. Знаходження кореляційної залежності між оцінкою якості перекладу яку дав експерт та оцінкою отриманою за розробленим алгоритмом.
3. Оцінити щільність зв'язку між відповідними ознаками за кореляційним відношенням, обчислити коефіцієнт детермінації.
4. Зробити висновок про адекватність оцінки за розробленим методом, оцінці перекладу експертом.

В результаті експерименту із 4 системами машинного перекладу мови, було отримано такі результати оцінки:

Таблиця 3.1 - Результати оцінювання систем МП за допомогою розробленого алгоритму(МЕ) та за допомогою експерта(НЕ)

| Система МП       | МЕ       | НЕ    |
|------------------|----------|-------|
| pragma           | 0.351481 | 3.977 |
| Google Translate | 0.344167 | 3.94  |
| Prompt           | 0.256204 | 3.858 |

Побудуємо гістограму отриманих результатів (для зручності значення оцінок НЕ розділимо на 10):

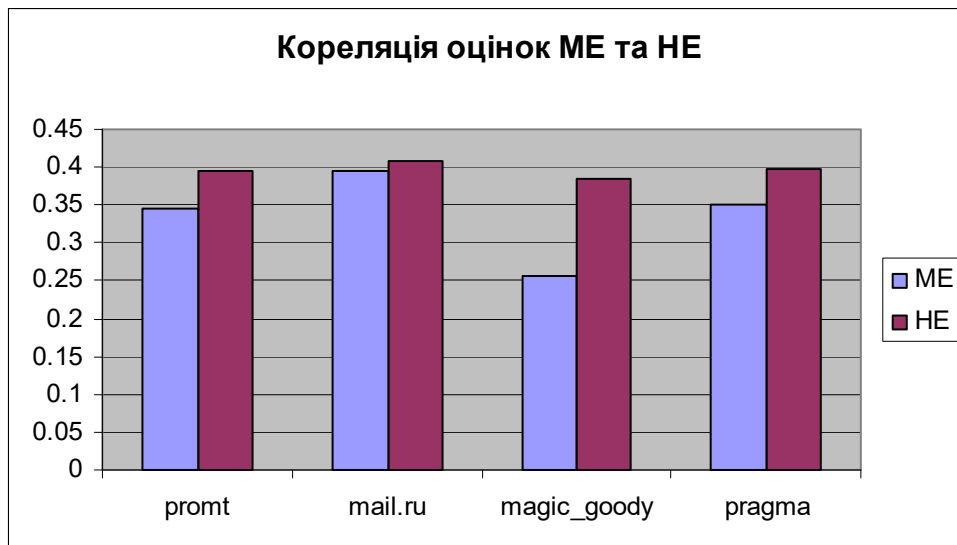


Рисунок 3.13 - Гістограма оцінок систем МП за допомогою розробленого алгоритму (МЕ) та за допомогою експерта (НЕ)

Знайдемо кореляційну залежність між оцінкою якості переклад, яку дав експерт(НЕ), та оцінкою отриманою за розробленим алгоритмом(МЕ). По емпіричним точкам з рисунку 3.16 видно, що залежність  $\overline{HE}_k = HE(ME)$ , доцільно шукати у вигляді лінійної функції:

$$\overline{HE}_k = a_1 ME + a_0. \quad (3.1)$$

Для знаходження параметрів функції використаємо систему нормальних рівнянь:

$$\begin{cases} a_1 \sum ME_i + na_0 = \sum HE_i, \\ a_1 \sum ME_i^2 + a_0 \sum ME_i = \sum HE_i \cdot ME_i, \end{cases} \quad (3.2)$$

Знайдемо середнє значення  $HE$  :

$$\overline{HE} = \frac{HE_i}{n}, \quad (3.3)$$

Тоді щільність зв'язку між  $HE$  та  $ME$  обчислюємо за допомогою коефіцієнта кореляції:

$$r_{HE/ME} = \sqrt{\frac{\sum (HE_k - \overline{HE})^2}{\sum (ME_k - \overline{ME})^2}}, \quad (3.4)$$

Проведемо розрахунки в середовищі MS Excel:

| Система МП                                                                                                                                                                                                                                                                                                                                 | ME       | HE     | ME <sup>2</sup> | HE*ME    | HEк(с)   | (HEк(с)-HEс) <sup>2</sup> | (HE-HEс) <sup>2</sup> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|--------|-----------------|----------|----------|---------------------------|-----------------------|
| promt                                                                                                                                                                                                                                                                                                                                      | 0.344167 | 3.94   | 0.118451        | 1.356017 | 3.973956 | 0.000120044               | 0.000529              |
| mail.ru                                                                                                                                                                                                                                                                                                                                    | 0.39537  | 4.077  | 0.156318        | 1.611925 | 4.050169 | 0.007598411               | 0.012996              |
| magic_goody                                                                                                                                                                                                                                                                                                                                | 0.256204 | 3.858  | 0.06564         | 0.988434 | 3.843031 | 0.014392554               | 0.011025              |
| pragma                                                                                                                                                                                                                                                                                                                                     | 0.351481 | 3.977  | 0.123539        | 1.397842 | 3.984844 | 0.000477159               | 0.000196              |
| сумма                                                                                                                                                                                                                                                                                                                                      | 1.347222 | 15.852 | 0.463948        | 5.354217 |          | 0.022588168               | 0.024746              |
| <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;"> <p>given</p> <p><math>1.347222n + 4a0 = 15.852</math></p> <p><math>0.463948n + 1.347222n0 = 5.354217</math></p> <p><math>\text{find}(a1, a0) \rightarrow \begin{cases} 1.488415607878782749 \\ 3.461693436980582636 \end{cases}</math></p> </div> |          |        |                 |          |          |                           |                       |
| a0=                                                                                                                                                                                                                                                                                                                                        | 3.461693 | HEс=   | 3.963           |          |          |                           |                       |
| a1=                                                                                                                                                                                                                                                                                                                                        | 1.488416 |        |                 |          |          |                           |                       |
| Обчислений коефіцієнт кореляції r =                                                                                                                                                                                                                                                                                                        |          |        |                 |          |          | 0.955406089               |                       |
| Коефіцієнт кореляції знайдено за допомогою функції КОРЕЛ() r =                                                                                                                                                                                                                                                                             |          |        |                 |          |          | 0.9553909                 |                       |

Рисунок 3.14 - Розрахунки коефіцієнта кореляції в середовищі MS Excel

Оскільки значення коефіцієнта кореляції близьке за абсолютним значенням до одиниці, то це свідчить про тісний зворотний зв'язок між оцінкою якості перекладу яку дав експерт та оцінкою отриманою за розробленим алгоритмом. Коефіцієнт детермінації  $D$  є підкореневим виразом формули (3.4). Цей коефіцієнт  $D \approx 0.912$  показує, що варіація фактора  $ME$  на 91% відбувається під впливом фактора  $HE$ , а на 9% під впливом неврахованих в моделі факторів.

Аналізуючи отримані результати можна зробити висновок про адекватність розробленої математичної моделі оцінювання систем машинного перекладу мови.

### 3.4 Інструкція користувачеві

Для початку роботи з програмою потрібно запустити виконуючий файл TextAnalyze.exe. Після цього на екрані з'явиться вікно, зовнішній вигляд якого представлений на рис. 3.5. Ви маєте змогу, обравши пункт Файл відкрити текст перекладений машиною → OpenText1, обравши Файл → OpenText2 відкрити текст перекладений людиною-експертом (ці файли можна створити самостійно). В меню текст представленому на рисунку 3.15, можна провести порівняння текстів(Analyze), відфільтрувати тест, або очистити його.

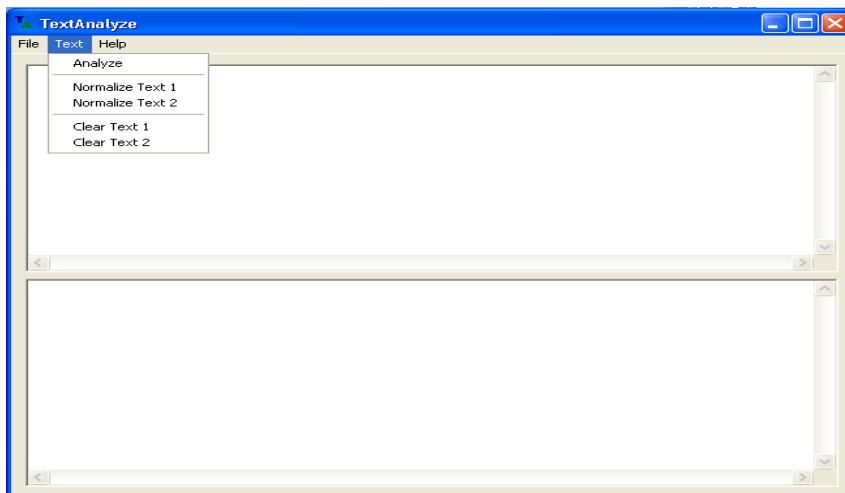


Рисунок 3.15 – Меню Text програми TextAnalyze.exe

Програму розроблено в середовищі Microsoft Visual Studio 2017. Мінімальна конфігурація комп'ютера для роботи програми Pentium 100 Mhz/ 32Mb ОЗУ/ 0.5 Mb HDD/ відеоадаптер 2 Mb.

### 3.5 Висновки

На основі розробленої моделі оцінювання систем МП розроблено програмне забезпечення, яке можна використовувати як автоматизованого дублера кваліфікованих експертів з перекладу, а також замість них у випадках, коли є потреба у швидких або частих оцінюваннях.

Розроблено методикау автоматизованого оцінювання систем машинного

перекладу мови на основі запропонованих методів, що дозволяє підвищити її адекватність шляхом врахування морфологічної і синтаксичної структури речень.

Запропоновано алгоритм порівняння систем автоматичного перекладу мови, що дозволяє розпізнавати закономірності та характеристики, притаманні тим чи іншим системам перекладу.

Автоматизовано процес оцінювання систем автоматичного перекладу мови, що дозволяє досліджувати вплив використовуваних алгоритмів перекладу на якість перекладу.

## 4 ЕКОНОМІЧНА ЧАСТИНА

### 4.1 Оцінювання комерційного потенціалу розробки

Метою проведення технологічного аудиту є оцінювання комерційного потенціалу розробки методів оцінки ефективності автоматичного перекладу мови в інтелектуальних системах автоматизації.

Для проведення технологічного аудиту було залучено 3-х незалежних експертів Вінницького національного технічного університету, кафедри комп'ютерних систем управління: Штовба С. Д., Ковалюк О. О., Ковтун В. В. За допомогою таблиці 4.1 за п'ятибальною шкалою використовуючи 12 критеріїв оцінки комерційного потенціалу розробки експерти надали свої оцінки.

Таблиця 4.1 – Рекомендовані критерії оцінювання комерційного потенціалу розробки та їх можлива бальна оцінка

| Критерії оцінювання та бали (за 5-ти бальною шкалою) |                                                                        |                                                                       |                                                             |                                                                       |                                                                        |
|------------------------------------------------------|------------------------------------------------------------------------|-----------------------------------------------------------------------|-------------------------------------------------------------|-----------------------------------------------------------------------|------------------------------------------------------------------------|
| Кри-терій                                            | 0                                                                      | 1                                                                     | 2                                                           | 3                                                                     | 4                                                                      |
| Технічна здійсненність концепції:                    |                                                                        |                                                                       |                                                             |                                                                       |                                                                        |
| 1                                                    | Достовірність концепції не підтверджена                                | Концепція підтверджена експертними висновками                         | Концепція підтверджена розрахунками                         | Концепція перевірена на практиці                                      | Перевірено роботоздатність продукту в реальних умовах                  |
| Ринкові переваги (недоліки):                         |                                                                        |                                                                       |                                                             |                                                                       |                                                                        |
| 2                                                    | Багато аналогів на малому ринку                                        | Мало аналогів на малому ринку                                         | Кілька аналогів на великому ринку                           | Один аналог на великому ринку                                         | Продукт не має аналогів на великому ринку                              |
| 3                                                    | Ціна продукту значно вища за ціни аналогів                             | Ціна продукту дещо вища за ціни аналогів                              | Ціна продукту приблизно дорівнює цінам аналогів             | Ціна продукту дещо нижче за ціни аналогів                             | Ціна продукту значно нижче за ціни аналогів                            |
| 4                                                    | Технічні та споживчі властивості продукту значно гірші, ніж в аналогів | Технічні та споживчі властивості продукту трохи гірші, ніж в аналогів | Технічні та споживчі властивості продукту на рівні аналогів | Технічні та споживчі властивості продукту трохи кращі, ніж в аналогів | Технічні та споживчі властивості продукту значно кращі, ніж в аналогів |



Продовження табл. 4.1

|                         |                                                                                                                              |                                                                                                                                      |                                                                                                                   |                                                                                           |                                                                                     |
|-------------------------|------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| 5                       | Експлуатаційні витрати значно вищі, ніж в аналогів                                                                           | Експлуатаційні витрати дещо вищі, ніж в аналогів                                                                                     | Експлуатаційні витрати на рівні експлуатаційних витрат аналогів                                                   | Експлуатаційні витрати трохи нижчі, ніж в аналогів                                        | Експлуатаційні витрати значно нижчі, ніж в аналогів                                 |
| Ринкові перспективи     |                                                                                                                              |                                                                                                                                      |                                                                                                                   |                                                                                           |                                                                                     |
| 6                       | Ринок малий і не має позитивної динаміки                                                                                     | Ринок малий, але має позитивну динаміку                                                                                              | Середній ринок з позитивною динамікою                                                                             | Великий стабільний ринок                                                                  | Великий ринок з позитивною динамікою                                                |
| 7                       | Активна конкуренція великих компаній на ринку                                                                                | Активна конкуренція                                                                                                                  | Помірна конкуренція                                                                                               | Незначна конкуренція                                                                      | Конкурентів немає                                                                   |
| Практична здійсненність |                                                                                                                              |                                                                                                                                      |                                                                                                                   |                                                                                           |                                                                                     |
| 8                       | Відсутні фахівці як з технічної, так і з комерційної реалізації ідеї                                                         | Необхідно наймати фахівців або витратити значні кошти та час на навчання наявних фахівців                                            | Необхідне незначне навчання фахівців та збільшення їх штату                                                       | Необхідне незначне навчання фахівців                                                      | Є фахівці з питань як з технічної, так і з комерційної реалізації ідеї              |
| 9                       | Потрібні значні фінансові ресурси, які відсутні. Джерела фінансування ідеї відсутні                                          | Потрібні незначні фінансові ресурси. Джерела фінансування відсутні                                                                   | Потрібні значні фінансові ресурси. Джерела фінансування є                                                         | Потрібні незначні фінансові ресурси. Джерела фінансування є                               | Не потребує додаткового фінансування                                                |
| 10                      | Необхідна розробка нових матеріалів                                                                                          | Потрібні матеріали, що використовуються у військово-промисловому комплексі                                                           | Потрібні дорогі матеріали                                                                                         | Потрібні досяжні та дешеві матеріали                                                      | Всі матеріали для реалізації ідеї відомі та давно використовуються у виробництві    |
| 11                      | Термін реалізації ідеї більший за 10 років                                                                                   | Термін реалізації ідеї більший за 5 років. Термін окупності інвестицій більше 10-ти років                                            | Термін реалізації ідеї від 3-х до 5-ти років. Термін окупності інвестицій більше 5-ти років                       | Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій від 3-х до 5-ти років | Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій менше 3-х років |
| 12                      | Необхідна розробка регламентних документів та отримання великої кількості дозвільних документів на виробництво та реалізацію | Необхідно отримання великої кількості дозвільних документів на виробництво та реалізацію продукту, що вимагає значних коштів та часу | Процедура отримання дозвільних документів для виробництва та реалізації продукту вимагає незначних коштів та часу | Необхідно тільки повідомлення відповідним органам про виробництво та реалізацію продукту  | Відсутні будь-які регламентні обмеження на виробництво та реалізацію продукту       |

Таблиця 4.2 – Рівні комерційного потенціалу розробки

| Середньоарифметична сума балів СБ, розрахована на основі висновків експертів | Рівень комерційного потенціалу розробки |
|------------------------------------------------------------------------------|-----------------------------------------|
| 0-10                                                                         | Низький                                 |
| 11-20                                                                        | Нижче середнього                        |
| 21-30                                                                        | Середній                                |
| 31-40                                                                        | Вище середнього                         |
| 41-48                                                                        | Високий                                 |

В таблиці 4.3 наведено результати оцінювання експертами комерційного потенціалу розробки.

Таблиця 4.3 – Результати оцінювання комерційного потенціалу розробки

| Критерії                                       | Прізвище, ініціали, посада експерта                                         |                     |                     |
|------------------------------------------------|-----------------------------------------------------------------------------|---------------------|---------------------|
|                                                | Штовба С.Д                                                                  | Ковалюк О.О.        | Ковтун В.В.         |
|                                                | Бали, виставлені експертами:                                                |                     |                     |
| 1                                              | 3                                                                           | 3                   | 2                   |
| 2                                              | 3                                                                           | 3                   | 3                   |
| 3                                              | 4                                                                           | 3                   | 3                   |
| 4                                              | 4                                                                           | 3                   | 3                   |
| 5                                              | 4                                                                           | 4                   | 4                   |
| 6                                              | 3                                                                           | 3                   | 2                   |
| 7                                              | 3                                                                           | 3                   | 4                   |
| 8                                              | 3                                                                           | 4                   | 4                   |
| 9                                              | 3                                                                           | 2                   | 3                   |
| 10                                             | 4                                                                           | 4                   | 4                   |
| 11                                             | 4                                                                           | 4                   | 4                   |
| 12                                             | 4                                                                           | 4                   | 4                   |
| Сума балів                                     | СБ <sub>1</sub> =42                                                         | СБ <sub>2</sub> =40 | СБ <sub>3</sub> =40 |
| Середньоарифметична сума балів $\overline{СБ}$ | $\overline{СБ} = \frac{\sum_{i=1}^3 СБ_i}{3} = \frac{42 + 40 + 40}{3} = 41$ |                     |                     |

Середньоарифметична сума балів, розрахована на основі висновків експертів склала 41, що згідно таблиці 4.2 вважається, що рівень комерційного потенціалу розробки є високим.

Результати роботи будуть впроваджені в навчальний процес кафедри КСУ, а також використовуються в наукових дослідженнях, що проводяться в рамках

держбюджетної тематики кафедральної науково-дослідної роботи № 52К4 “Методи прийняття рішень на основі принципів обчислювального інтелекту”.

#### 4.2 Прогнозування витрат на виконання науково-дослідної роботи

Проведемо прогнозування витрат на виконання робіт у три етапи:

*1-й етап:* Розрахунок витрат, які безпосередньо стосуються виконавців даного розділу НДР.

*2-й етап:* Розрахунок загальних витрат на виконання НДР;

*3-й етап:* Прогнозування загальних витрат на виконання та впровадження НДР.

Проведемо перший етап прогнозування.

Основна заробітна плата кожного із дослідників  $Z_0$ , якщо вони працюють в наукових установах бюджетної сфери визначається за формулою:

$$Z_0 = \frac{M}{T_p} * t \text{ (грн)} \quad (4.1)$$

де  $M$  – місячний посадовий оклад конкретного розробника (інженера, дослідника, науковця тощо), грн.;

$T_p$  – число робочих днів в місяці; приблизно  $T_p \approx 21...23$  дні;

$t$  – число робочих днів роботи дослідника.

Дану розробку буде проводити інженер-консультант, величина окладу буде становити 7000 грн. на місяць. Кількість робочих днів у місяці складає 22, а кількість робочих днів дослідника складає 45. Зведемо сумарні розрахунки до таблиця 4.4.

#### 2. Розрахунок додаткової заробітної плати робітників

Додаткова заробітна плата  $Z_d$  всіх розробників та робітників, які приймали участь в розробці нового технічного рішення розраховується як 10 - 12 % від основної заробітної плати робітників.

Таблиця 4.4 – Заробітна плата дослідника в науковій установі бюджетної сфери

| Найменування посади | Місячний посадовий оклад, грн. | Оплата за робочий день, грн. | Число днів роботи | Витрати на заробітну плату грн. |
|---------------------|--------------------------------|------------------------------|-------------------|---------------------------------|
| Керівник проекту    | 10000                          | 454.5                        | 5                 | 2273                            |
| Інженер-консультант | 7000                           | 318.2                        | 45                | 14318                           |
| Всього              |                                |                              |                   | 16591                           |

На даному підприємстві додаткова заробітна плата начисляється в розмірі 10% від основної заробітної плати.

$$Z_{д} = 0,11 * 16591 = 1825 \text{ (грн)} \quad (4.2)$$

3. Нарахування на заробітну плату  $H_{зп}$  дослідників та робітників, які брали участь у виконанні даного етапу роботи, розраховуються за формулою (4.3):

$$H_{зп} = (Z_0 + Z_{д}) * \frac{\beta}{100} \text{ (грн)} \quad (4.3)$$

де  $Z_0$  – основна заробітна плата розробників, грн.;

$Z_{д}$  – додаткова заробітна плата всіх розробників та робітників, грн.;

$\beta$  – ставка єдиного внеску на загальнообов'язкове державне соціальне страхування, % .

Дана діяльність відноситься до бюджетної сфери, тому ставка єдиного внеску на загальнообов'язкове державне соціальне страхування буде складати 22%, тоді:

$$H_{зп} = (16591 + 1825) * \frac{22}{100} = 4051,5 \text{ (грн)}$$

4. Амортизація обладнання, комп'ютерів та приміщень, які використовувались під час виконання даного етапу роботи

Дані відрахування розраховують по кожному виду обладнання, приміщенням тощо.

$$A = \frac{Ц \cdot T}{T_{кор} \cdot 12} \text{ [грн]}, \quad (4.4)$$

де Ц – балансова вартість даного виду обладнання (приміщень), грн.;

$T_{кор}$  – час користування;

T – термін використання обладнання (приміщень), цілі місяці.

Згідно пункту 137.3.3 Податкового кодекса амортизація нараховується на основні засоби вартістю понад 2500 грн. В нашому випадку для написання магістерської роботи використовувався персональний комп'ютер вартістю 11000 грн.

$$A = \frac{11000 \cdot 2}{2 \cdot 12} = 916,67$$

5. Норма витрат матеріалу – це плановий показник, який визначає максимально допустимі затрати відповідних ресурсів на виробництво одиниці продукції в умовах певного рівня техніки і організації виробництва.

Витрати на матеріали M, що були використані під час виконання даного етапу роботи, розраховуються по кожному виду матеріалів за формулою:

$$M = \sum_1^n H_i \cdot Ц_i \cdot K_i - \sum_1^n V_i \cdot Ц_v \quad \text{грн.}, \quad (4.5)$$

де  $H_i$  – витрати матеріалу  $i$ -го найменування, кг;

$Ц_i$  – вартість матеріалу  $i$ -го найменування, грн./кг.;

$K_i$  – коефіцієнт транспортних витрат,  $K_i = (1,1 \dots 1,15)$ ;

$V_i$  – маса відходів матеріалу  $i$ -го найменування, кг;

$Ц_v$  – ціна відходів матеріалу  $i$ -го найменування, грн/кг;

n – кількість видів матеріалів.

Інформацію про використані матеріали подамо у вигляді табл. 4.5.

Таблиця 4.5– Матеріали, що використані на розробку

| Найменування матеріалу                    | Ціна за одиницю, грн. | Витрачено | Вартість витраченого матеріалу, грн. |
|-------------------------------------------|-----------------------|-----------|--------------------------------------|
| Папір                                     | 85                    | 1         | 85                                   |
| Ручка                                     | 11                    | 1         | 11                                   |
| Флешка                                    | 110                   | 1         | 110                                  |
| CD-диск                                   | 12                    | 1         | 12                                   |
| Всього                                    |                       |           | 218                                  |
| З врахуванням коефіцієнта транспортування |                       |           | 239,8                                |

6. До статті «Паливо та енергія на технологічні цілі» відносяться витрати на всі види палива й енергії, що безпосередньо використовуються у процесі виробництва продукції. У даному випадку будемо враховувати лише витрати на електроенергію, яку споживає комп'ютер:

$$B_e = B \cdot P \cdot \Phi \cdot K_p, \quad (4.6)$$

де  $B$  – вартість 1 кВт енергії, грн.  $B = 8,44$  грн/кВт\*год;

$P$  – установлена потужність обладнання, кВт.  $P = 500$  Вт або  $P = 0,5$  кВт;

$\Phi$  – фактична кількість годин роботи обладнання, год.  $\Phi = 180$  год.;

$K_p$  – коефіцієнт використання потужності,  $K_p = 0,65$ .

$$B_e = 8,44 \cdot 0,5 \cdot 180 \cdot 0,65 = 493,74 \text{ (грн)}.$$

Інші витрати  $B_{in}$  охоплюють: витрати на управління організацією, оплата службових відряджень, витрати на утримання, ремонт та експлуатацію основних засобів, витрати на опалення, освітлення, водопостачання, охорону праці тощо. Інші витрати  $B_{in}$  можна прийняти як (100...300)% від суми основної заробітної плати розробників та робітників, які виконували дану МКНР, тобто:

$$B_{in} = (1..3) \cdot (3 + 3_p). \quad (4.7)$$

$$B_{in} = 1 \cdot 16591 = 16591 (\text{грн.})$$

Сума всіх попередніх статей витрат дає витрати, які безпосередньо стосуються даного розділу МКНР

$$B = 16591 + 1825 + 4051,5 + 916,67 + 239,8 + 493,74 + 16591 = 40708,5 (\text{грн.})$$

Загальна вартість всієї МКНР визначається за формулою:

$$B_{заг} = \frac{B}{\alpha} \quad (4.8)$$

$$B_{заг} = \frac{40708,5}{1} = 40708,5 (\text{грн.})$$

Прогнозування загальних втрат ЗВ на виконання та впровадження результатів виконаної МКНР здійснюється за формулою:

$$ЗВ = \frac{B}{\beta}, \quad (4.9)$$

де  $\beta$  – коефіцієнт, який характеризує стадію виконання даної НДР.

Оскільки, робота знаходиться на стадії науково-дослідних робіт, то коефіцієнт  $\beta = 0,9$ .

Звідси:

$$ЗВ = \frac{40708,5}{0,9} = 45232 (\text{грн.}).$$

#### 4.3 Оцінка внеску НДР

Для цього спочатку потрібно визначити коефіцієнт наукової значимості отриманих результатів НДР, а потім підрахувати внесок в досягнення цих результатів.

Коефіцієнт наукової значимості результатів проведеної науково-дослідної роботи  $K_{ЗН}$  можна підрахувати за формулою:

$$K_{ЗН} = \frac{\sum_{i=1}^3 b_i \cdot d_i}{\sum_{i=1}^3 b_{max} \cdot d_i} \quad (4.10)$$

де  $b_i$  – значимість отриманих результатів:  $b_1$  – ступінь наукової новизни,  $b_2$  – рівень теоретичної обґрунтованості,  $b_3$  – ступінь експериментальної перевірки результатів. Бальна оцінка отриманих результатів наведена в таблиці 4.8. Максимальне значення отриманих за кожною характеристикою результатів можна прийняти в межах 7...10 балів;

$d_i$  – питома вага кожної характеристики, значення якої наведені в таблиці 4.8;

3 – кількість характеристик, за якими була зроблена оцінка результатів науково-дослідної роботи.

Таблиця 4.7 – Показники для оцінювання наукової значимості результатів виконаної НДР

| Характеристик<br>и | Питома вага<br>характерист<br>ики $d_i$ | Бальна оцінка характеристики                                                                  |                                                                                              |                                                                                                                                           |
|--------------------|-----------------------------------------|-----------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
|                    |                                         | Ступінь новизни $b_1$                                                                         | Рівень теоретичної<br>обґрунтованості $b_2$                                                  | Ступінь<br>експериментальної<br>перевірки результатів<br>$b_3$                                                                            |
|                    |                                         | 1                                                                                             | 2-5                                                                                          | 7-10                                                                                                                                      |
| $b_1$              | 0,500                                   | Часткове<br>удосконалення виробів,<br>технологій, матеріалів,<br>програмного продукту<br>тощо | Суттєве удосконалення<br>виробів, технологій,<br>матеріалів,<br>програмного продукту<br>тощо | Нові напрямки в<br>розробці виробів,<br>технологій, матеріалів,<br>програмного продукту<br>тощо. Створення<br>принципово нової<br>техніки |
| $b_2$              | 0,333                                   | Позитивне рішення на<br>основі зроблених<br>узагальнень                                       | Установлення<br>залежності, які<br>використовувались в<br>інших випадках                     | Відкриття нових шляхів<br>рішення задачі                                                                                                  |
| $b_3$              | 0,167                                   | Експериментальна<br>перевірка не робилась                                                     | Результати<br>перевірялись на<br>невеликій кількості<br>даних                                | Результати перевірені<br>на великій кількості<br>даних                                                                                    |

В ході виконання даної магістерської кваліфікаційної роботи було розроблено метод оцінки ефективності автоматичного перекладу мови в інтелектуальних системах автоматизації.

Тоді згідно таблиці 4.8  $b_1 = 5$ . Отримані результати підтвердили залежності та характеристики, отримані раніше для окремих випадків рішення даної задачі, тому



$b_2 = 6$ . Оскільки результати перевірялись на отриманих попередньо експериментальних даних, приймаємо  $b_3 = 8$ . Тоді коефіцієнт наукової значимості результатів буде наступним

$$K_{ZH} = \frac{5 \cdot 0,5 + 6 \cdot 0,333 + 8 \cdot 0,167}{10 \cdot 0,5 + 10 \cdot 0,333 + 10 \cdot 0,167} = 0,58.$$

Зважаючи на малий час, відведений на проведення дослідження, отримане значення коефіцієнта наукової значимості можна вважати цілком задовільним.

Внесок дипломника в досягнення отриманих результатів НДР можна розрахувати за формулою:

$$V = \frac{\sum_{i=1}^n k_{ТВі} \cdot Z_i}{\sum_{i=1}^n k_{ТВі} \cdot Z_i} \quad (4.11)$$

де  $k_{ТВі}$  – коефіцієнт творчої участі кожного виконавця НДР, який оцінюється таким чином: проведення досліджень – 3 бали, робоче проектування – 1,5 бали, освоєння – 1,0 балів. якщо виконавець приймав участь в декількох видах робіт, то береться сума відповідних балів;

$Z_i$  – заробітна плата кожного виконавця НДР, грн.;

$n$  – кількість всіх виконавців науково-дослідної роботи: наукові керівники, відповідальні виконавці, спеціалісти, наукові співробітники, робітники різних професій тощо.

$$V = \frac{3 \cdot 10000}{3 \cdot 10000 + 7000 \cdot 1,5} = 0,74$$

#### 4.4 Висновок

В даному розділі було оцінено комерційний потенціал розробки озробка методу оцінки ефективності автоматичного перекладу мови в інтелектуальних системах автоматизації. Комерційний потенціал розробки виявився високому рівні.

Прогнозування витрат на виконання науково-дослідної роботи по кожній з статей витрат складе 40708,5 грн. Загальна ж величина витрат на виконання та впровадження результатів даної НДР буде складати 45232 грн.

Отримане значення коефіцієнта наукової значимості ( $K_{3H} = 0,58$ ) можна вважати цілком задовільним, оскільки час відведений на проведення дослідження незначний. Особистий внесок дипломника в наукове дослідження становить 0,74, що являється нормальним значенням.

## ВИСНОВКИ

В магістерській кваліфікаційній роботі досліджуються методи оцінки ефективності систем автоматичного перекладу мови.

В першому розділі проведено аналіз сучасного стану проблеми машинного перекладу мови: наведено коротку історію становлення машинного перекладу, проведено класифікацію систем МП, розглянуто проблеми побудови систем МП, зроблено постановку проблеми для подальших досліджень. Було зазначено, що однією з актуальних невирішених задач на сьогодні в області машинного перекладу мови є розробка методів оцінки якості роботи систем МП, а отже необхідно розробити адекватні методи оцінки, які були б на одному рівні з оцінками експертів.

В другому розділі проведено математичне обґрунтування методів оцінки систем автоматичного перекладу мови: приведено аналіз існуючих алгоритмів побудови систем МП, розглянуто системи статистичного декодування та метод n-грам, основні принципи методу n-грам, модифікований метод n-грам із зміненою точністю, метод оцінки ефективності систем МП на основі інформаційно-статистичного критерію і моделі потенціальної системи. Запропоновано використання в синтаксичній діаграмі морфем замість слів та залучення семантичної інформації. Розроблено новий метод оцінки ефективності систем машинного перекладу, що ґрунтується на інформаційно-статистичному критерії і моделі потенціальної системи, який відрізняється тим, що поряд із статистичними характеристиками мови, враховує і її синтаксичну структуру, що дає змогу оцінити не тільки статистичну, а й синтаксичну інформацію про мову.

В третьому розділі на основі запропонованих методів і моделей розроблено аналітичні, алгоритмічні і програмні засоби:

- методику автоматизованого оцінювання систем машинного перекладу мови на основі запропонованих методів, що дозволяє підвищити її адекватність шляхом врахування морфологічної і синтаксичної структури речень;

- програмне забезпечення для реалізації розроблених математичних моделей, що дозволяє використовувати його як автоматизованого дублера кваліфікованих експертів з перекладу, а також замість них у випадках, коли є потреба у швидких або частих оцінюваннях;

- запропоновано алгоритм порівняння систем автоматичного перекладу мови, що дозволяє розпізнавати закономірності та характеристики, притаманні тим чи іншим системам перекладу.

- проведено експериментальні дослідження, в результаті яких було одержано результати оцінки систем автоматичного перекладу мови, які корельовані із оцінками експертів. розроблено алгоритми для реалізації методу порівняння кількості збігів-грам, наведено контрольний приклад та аналіз отриманих результатів, розроблено супроводжувальну документацію.

Основні положення й результати досліджень доповідалися й обговорювалися на форумі молодих науковців ВНТУ (м. Вінниця, 2019). Результати роботи впроваджені в навчальний процес кафедри КСУ, а також використовуються в наукових дослідженнях, що проводяться в рамках держбюджетної тематики кафедральної науково-дослідної роботи № 52К4 “Методи прийняття рішень на основі принципів обчислювального інтелекту”.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Papineni K., Roukos S., Ward T. BLEU: a Method for Evaluation of Mashine Translation // Proceedings of the 40-th Annual Meeting of the Assoc. Comp. Linguist. (ACL). – Philadelphia, July, 2002. – P. 311-318.
2. Akiba Y., K Imamura and E. Sumita. 2001. Using multiple edit distances to automatic rank machine translation output // MT Summit VIII. P. 15–20.
3. Akiba Y., E. Sumita, H. Nakaiwa, S. Yamamoto and H.G. Okuno Experimental Comparison of MT Evaluation Methods: RED vs. BLEU. // Proc. MT Summit IX,- 2003.
4. Brew C., Thompson H. 1994. Automatic Evaluation of Computer Generated Text. ARPA/ISTO Workshop on Human Language Technology, 1994. P. 104-109
5. MT Summit IX. <http://www.amtaweb.org/summit/MTSummit/FinalPapers/55-Akiba-final.pdf>.
6. Babych, B; Hartley, A.; Atwell, E. Statistical Modelling of MT output corpora for Information Extraction // Proceedings of the Corpus Linguistics 2003 conference, edited by Dawn Archer, Paul Rayson, Andrew Wilson and Tony McEnery. Lancaster University (UK), 28 - 31 March 2003. P. 62-70.
7. Salton, G. and M.E. Lesk.. Computer evaluation of indexing and text processing. Journal of the ACM, 15(1), - 1968. - P.8-36
8. Papineni K, Roukos S, Ward T, Zhu W-J 2001 Bleu: a method for automatic evaluation of machine translation // IBM research report RC22176 (W0109-022) September 17, 2001
9. Rajman, M. and T. Hartley. 2001. Automatically predicting MT systems ranking compatible with Fluency, Adequacy and Informativeness scores // Proceedings of the 4<sup>th</sup> ISLE Workshop on MT Evaluation, MT Summit VIII. Santiago de Compostela, September 2001. pp. 29-34.
10. Penrose, R. The Emperors New Mind // Oxford University Press. - 1989. – P. 15.
11. White, J., T. O’Connell and F. O’Mara. 1994. The ARPA MT evaluation methodologies: evolution, lessons and future approaches // Proceedings of the 1<sup>st</sup> Conference of the Association for Machine Translation in the Americas. Columbia, MD, October 1994. pp. 193-205.

12. Винокуров А. А. Метод оценки качества перевода. – Режим доступа: <http://www.ito.su/1998/2/VINIKU.html>. - Заголовок з екрану.
13. Винокуров А.А., Чуканов В.О. Новый метод оценки машинного перевода // Информационные технологии и системы. Hardware Software Security. Тенденции и перспективы - Сборник статей: М., Международная академия информатизации, 1997г
14. Гавришук В.Л. Биков М.М. Аналіз методів оцінки ефективності систем машинного перекладу мови // Тези студентських доповідей.- Вінниця: ВНТУ.- 2006.
15. Биков М.М., Балховський Д.Є. .Лінгвістико-математичний метод опису символів // Матеріали IV Міжнародної конференції студентів і молодих вчених. – К.: НАУ, 2004. – С.26
16. Trier S. D., Jain A.K., Taxt T. Feature extraction methods for character recognition – A Survey. // Pattern recognition, 1996. – Vol. 29, No. 4. – P. 641-642.
17. Федоров А.В. Основы общей теории перевода. - М.: Высшая школа, 1968. – 256 с.
18. Нагао М., Цудзии Д., Накамура Д. Машинный перевод с японского языка на английский // ТИИЭР. – 1986. – Т.74, №7. – с.112-133.
19. Обработка естественных языков / Под ред. Кулагиной Т. – М.: Знание, 1989. – 214 с.
20. Слокум Дж. Перспективы развития машинного перевода – мнение американских специалистов // ТИИЭР. – 1986. – Т.74, №7. – с. 77-86.
21. Комиссаров В.Н. Теория перевода (лингвистические аспекты). – Режим доступа: [http://www.classes.ru/grammar/43.Teoriya\\_perevoda\\_Lingvicticheskiye\\_aspekty/extfile/help2/html/unnamed\\_60.html](http://www.classes.ru/grammar/43.Teoriya_perevoda_Lingvicticheskiye_aspekty/extfile/help2/html/unnamed_60.html). - Заголовок з екрану.
22. Бельская И.К. Язык человека и машина. – М.: Знание, 1969. – 138 с.
23. Марчук Ю.Н. Проблемы машинного перевода. – М.: Знание, 1983. – 175 с.
24. Нелюбин Л.Л. Компьютерная лингвистика и машинный перевод. - М.: Наука, 1991. – 254 с.
25. Беляева Л.Н., Откупщикова М.И. Автоматический (машинный) перевод. – В сб.: Прикладное языкознание. С.-Пб., - 1996. – 143 с.

26. Прокопович С.С. Адекватный перевод или интерпретация текста? // Тетради переводчика. - 1980. - Вып. 17. - С. 37-48.
27. Семенюк Е.В. Проблема стилизации и перевод: (на материале английской и американской художественной литературы): Автореф. дис. - М., 2003.- 26 с.
28. Смирнов Л.Н. Отражение взаимодействия культур в переводе // Встречи этнических культур в зеркале языка: в сопоставительном лингвокультурном аспекте. - М.: Наука, 2002. - С. 378-394.
29. Солдатова Л.А. Лингвистические и экстралингвистические основы вариативности перевода: Автореф. дис. - М., 2002. - 24 с.
30. Журнал литературных переводов с английского языка и на английский. – Режим доступа: <http://www.uiowa.edu/~xchanges/index.htm>. - Заголовок з экрану.
31. The Translator's Home Companion // Ресурсы для переводчиков: новости в области перевода; образование в области перевода; конференции и семинары; инструментарий переводчика; глоссарии; автоматические переводчики и мн. др. – Режим доступа: <http://www.lai.com/companion.html>. - Заголовок з экрану.
32. Translation. Net Переводческие организации, клавиатуры на разных языках, ссылки на переводческие ресурсы. – Режим доступа: <http://www.translation.net>. - Заголовок з экрану.
33. Webtranslators.com Дискуссионные форумы и чаты по интересам специалистов по языкам, бесплатные электронные адреса с возможностью доступа через web, глоссарии и словари, переводческие организации, новости, ссылки на информацию о конференциях, образовании в области перевода и др. – Режим доступа: <http://www.webtranslators.com>. - Заголовок з экрану.
34. Peter Sandrini's Translation Resources Один из наиболее исчерпывающих источников по переводческим ресурсам. – Режим доступа: <http://homepage.uibk.ac.at/~c61302/tranlink.html>. - Заголовок з экрану.
35. ТрансИнтер Все о переводах, все для перевода Новости; форумы; ссылки на переводческие ресурсы; поиск информации в лучших онлайн-словарях, энциклопедиях, поисковых системах, новостях и т.п.; статьи. – Режим доступа: <http://www.transinter.ru/>. - Заголовок з экрану.

36. Translation Journal A Publication for Translators by Translators about Translators and Translation. – Режим доступу до журналу: <http://accurapid.com/journal/>.
37. New Voices in Translation Studies Электронный журнал по проблемам перевода, издаваемый IATIS (International Association for Translation & Intercultural Studies) и CTTS (Centre for Translation and Textual Studies). Выходит 1 раз в год. (пока представлен один номер за 2005 г.). – Режим доступу до журналу: <http://www.iatis.org/newvoices/index.htm>.
38. Бикова К.М., Гаврищук В.Л., Биков М.М. Аналіз сучасного стану проблеми машинного перекладу мови // Матеріали IV міжнародної конференції студентів та молодих вчених. - С.27. - Київ: Національний авіаційний університет.- 2004.- 416 с.
39. Бикова К.М., Гаврищук В.Л. Сучасні проблеми машинного перекладу мови // Тези студентських доповідей.- С.9.- 204: ВНТУ, - 2004.- 237 с.
40. Гаврищук В.Л. Биков М.М. Підвищення ефективності методів оцінки систем автоматичного перекладу мови // Тези студентських доповідей.- Вінниця: ВНТУ.- 2007.
41. Гаврищук В.Л. Биков М.М. Методи оцінки ефективності систем автоматичного перекладу мови // V міжнародна науково-практична конференція. – Хмельницький: ХНУ, - 2007.
42. Комиссаров, В.Н. Современное переводоведение: Учеб. пособие В.Н.Комиссаров. - М.: ЭТС,2002.- 420 с.
43. Говорун М. Машинный перевод в XX веке // Мир Internet.-2001. - № 2. - С.16-26
44. Ляпунов А. А., Кулагина О. С. Использование вычислительных машин для перевода с одного языка на другой // Природа. – 1955, № 8. – с. 37-33.
45. Кузнецов П. С., Ляпунов А. А., Реформатский А. А. Основные проблемы машинного перевода // Вопросы языкознания. – 1956, № 5. – с. 59-73.
46. Панов Д. Ю., Ляпунов А. А., Мухин И. С. Автоматизация перевода с одного языка на другой // Сессия по научным проблемам автоматизации производства. - М., Изд. АН СССР, 1956. – с. 78-94.
47. Беляева Л.Н. Автоматический (машинный) перевод / Л.Н. Беляева, М.И. Откупщикова // Прикладное языкознание. - СПб., 1996. - С. 360-388.



48. Переводчики и словари PROMT. – Режим доступа: <http://www.translate.ru>. - Заголовок з екрану.
49. Беспалова Н.П. Перевод и реферирование общественно-политических текстов // Английский язык: Учебное пособие. - 4. изд., перераб. и испр.- М.: Изд-во Российского университета дружбы народов, 2003.- 124 с.
50. Дмитрищенко Д.В. Переводческие трансформации для передачи аффективных лексических элементов английского научно-технического текста // Прагматические аспекты языкознания. - М., 2001. - С. 54-66.
51. Доброгост Л.А. Автоматическая обработка текста в системах машинного перевода // Перевод и переводоведение. -Уфа, 1997. - С. 24-29.
52. Нелюбин Л.Л. Компьютерная лингвистика и машинный перевод: Метод. пособие. - М., 1991. - 151 с.
53. Степанов, М. Ф. Машинный перевод и общение на естественном языке : учеб. пособие / М. Ф. Степанов ; М-во образования Рос. Федерации, Саратов. гос. техн. ун-т. - Саратов : СГТУ, 2000. - 96 с.
54. Гальченко, О. Н. Система интерактивного машинного перевода как воспроизводящая инженерно-лингвистическая модель : автореф. дис. ... канд. филол. наук / О.Н.Гальченко ; Воен. ин-т. - М., 1989. - 23 с. : схем. - Библиогр.: с. 22-23
55. Чужакин А.П., Палажченко П.Р. Мир перевода, или Вечный поиск взаимопонимания. – М.: Валент, 1999. - с. 33-37.
56. Шатков Г.В. Стилистический анализ текста при переводе // Тетради переводчика. - 1979. - Вып. 16. - С. 40-50.
57. Шафиров С.Г. Лексико-семантическая типология соматизмов в английском языке в соответствии с французским // Словообразование и лексические системы в разных языках. Вып.1 - Уфа, 1994. - С. 147-153.
58. Галеева Н.Л. Динамические факторы перевода // Языковые подсистемы: стабильность и движение. - Тверь, 2002. - С. 30-40.
59. Система перевода текста Magic Goody для Windows. Руководство пользователя. — С.-Петербург, фирма "ПРОМТ", 1999. – 260 с.

60. Система перевода текста PROMT Internet. Руководство пользователя. - С.-Петербург, фирма "ПРОМТ", 1999. – 262 с.
61. Сайт компанії Промт. – Режим доступу: <http://www.promt.ru>.
62. L.Visson. From Russian Into English. Ardis, 1991.
63. Babych B., Hartley A. (2004). Extending the BLEU MT Evaluation Method with Frequency Weightings. In ACL 2004 Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, Barcelona, July 2004, pp. 622-629.
64. Babych B., Hartley A., Elliott, D. (2005). Estimating the predictive power of n-gram evaluation metrics across languages and text types. In Proceedings of MT Summit X, pp. 12-16 September, 2005, Phuket, Thailand, pp 412-418.
65. Babych, B. Weighted N-gram model for evaluating Machine Translation output. In Proceedings of the 7<sup>th</sup> Annual Colloquium for the UK Special Interest Group for Computational Linguistics. University of Birmingham, 6-7 January, 2004. pp. 15-22.
66. University of Leeds // School of Computing, FACULTY OF ENGINEERING. – Режим доступу: <http://www.comp.leeds.ac.uk/bogdan/ltv-mt-eval.html>. - Заголовок з екрану.
67. Levinson S., Rabiner L., Sondhi M. An Introduction to the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition // Bell Systems Technical Journal, 1983. - Vol.62, No.4. – P. 1035 -1074.
68. Creutz M., Lagus K. Unsupervised Morpheme Segmentation and Morphology Induction from Text Corpora Using Morfessor 1.0. – Режим доступу: <http://www.cis.hut.fi/mcreutz/papers/Creutz04tr.pdf>. - Заголовок з екрану.
69. Биков М.М., Балховський Д.Є. Лінгвістико-математичний метод опису символів. - Кн.:Матеріали IV Міжнародної конференції студентів і молодих вчених. – К.: НАУ, 2004. – С.26
70. Быков Н.М., Агеев А.С. Модель потенциальной системы для распознавания речи в СЧМ // Исследование и проектирование систем “человек-машина”. – Киев: ИК АНН УССР, 1986. – с. 57-61.
71. Бром Н.С., Галунов В.И. Система распознавания речи. – В кн.: Речевое общение в автоматизированных системах. – М.: Наука, 1975, с. 5-13.
72. Методы автоматического распознавания речи // Под ред. У. Ли. – М.: Мир, 1983. – 716 с.

73. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов // Пер. с англ. М.: Радио и связь, 1981. – 496 с.
74. Сапожков М.А. Речевой сигнал в кибернетике и связи. – М.: Связьиздат, 1963. – 452 с.
75. Фант Г. Акустическая теория речеобразования. – М.: Наука, 1964. – 284 с.
76. Харкевич А.А. Спектры и анализ. – М.: Физматиз, 1962. – 320 с.
77. Маркел Д.Д., Грей А.Х. Линейное предсказание речи. – М.: Связь, 1980. – 380 с.
78. Бондарко Л.В., Загоруйко Н.Г. и др. Модель восприятия речи человеком. – Новосибирск: Наука, 1968. – 58 с.
79. С.П. Иглин Теория вероятностей и математическая статистика на базе MATLAB. Харьков: НТУ "ХПИ", 2006,-612 стр.
80. Кузьмин И.В. Оценка эффективности и оптимизация АСКУ // Советское радио. - 1971.-296 с.
81. Биков М.М., Грищук Т.В. Розробка методів оцінки ефективності автоматизованих систем розпізнавання мови // Вісник Технологічного університету Поділля – Хмельницький, ТУП, 2003. - №3, том 1 – с. 122-125.

## ДОДАТКИ

Додаток А  
(обов'язковий)  
ВНТУ

ТЕХНІЧНЕ ЗАВДАННЯ  
на виконання магістерської кваліфікаційної роботи  
«Розробка методів оцінки ефективності автоматичного перекладу мови в  
інтелектуальних системах автоматизації»

Студент групи 2АКІТ – 18м Власюк М. В.

“ \_\_\_\_\_ ” \_\_\_\_\_ 2019р.

Керівник к. т. н., доцент Биков М. М.

“ \_\_\_\_\_ ” \_\_\_\_\_ 2019р.

Вінниця 2019

## 1. Назва та галузь застосування

1.1. Назва – Розробка методів оцінки ефективності автоматичного перекладу мови в інтелектуальних системах автоматизації.

1.2. Галузь застосування – системи управління і автоматики, інформаційні технології.

## 2. Підстава для проведення розробки.

Тема магістерської дипломної роботи затверджена наказом по ВНТУ № \_\_\_\_ від “ \_\_\_\_ ” 09 2019 р.

## 3. Мета та призначення розробки.

Метою дисертаційної роботи є підвищення ефективності методів оцінювання систем автоматичного перекладу.

## 4. Вихідні дані для проведення розробки.

Магістерська кваліфікаційна робота виконується вперше. В ході проведення розробки повинні використовуватись такі документи:

1. Papineni K., Roukos S., Ward T. BLEU: a Method for Evaluation of Mashine Translation // Proceedings of the 40-th Annual Meeting of the Assoc. Comp. Linguist. (ACL). – Philadelphia, July, 2002. – P. 311-318.
2. Выков N.M., Kuzmin I.V., Yakovenko A.I. Development of effective strategy of pattern recognition // Proceedings of SPIE, 2000, Vol.4425, pp. 76– 82.
3. Слокум Дж. Перспективы развития машинного перевода – мнение американских специалистов // ТИИЭР. – 1986. – Т.74, №7. – с. 77-86.
4. Гаврищук В.Л. Биков М.М. Методи оцінки ефективності систем автоматичного перекладу мови // V міжнародна науково-практична конференція. – Хмельницький: ХНУ, - 2007. – С. 25 – 26.

## 5. Вимоги до розробки.

### 5.1. Перелік головних функцій:

- завантаження вихідного тексту;
- завантаження перекладеного вручну тексту;
- завантаження перекладеного машиною тексту;
- обчислення співпадінь методом зважених N-грам;

- обчислення оцінки якості машинного перекладу згідно з запропонованим методом.

## 5.2. Основні технічні вимоги до розробки.

### 5.2.1. Вимоги до програмної платформи:

- WINDOWS 7\8;
- Matlab 7.0;
- Microsoft Visual Studio.

### 5.2.2. Умови експлуатації системи:

- робота на стандартних ПЕОМ в приміщеннях зі стандартними умовами;
- можливість цілодобового функціонування системи;
- текст програмного забезпечення системи є цілком закритим.

## 6. Економічні показники

До економічних показників входять:

- термін окупності не більше \_\_\_\_\_ років;
- розмір капіталовкладень не більше \_\_\_\_\_ грн.;
- економічний ефект не менше \_\_\_\_\_ грн.;
- рентабельність не менше \_\_\_\_\_ %;
- інші економічні переваги у порівнянні з аналогами.

## 7. Стадії та етапи розробки.

### 7.1 Пояснювальна записка:

|   |                                                            |                      |
|---|------------------------------------------------------------|----------------------|
| 1 | Техніко-економічне обґрунтування доцільності розробки      | 02.10.2019 р.        |
| 2 | Постановка задачі і розробка технічного завдання           | 04.10.2019 р.        |
| 3 | Розробка моделі обробки текстових документів               | 20.10.2019 р.        |
| 4 | Розробка програми реалізації моделі і машинний експеримент | 15.11.2019 р.        |
| 5 | Підготовка економічної частини                             | 22.11.2019 р.        |
| 6 | <u>Оформлення пояснювальної записки,</u>                   | <u>30.11.2019 р.</u> |

|    |                                   |               |
|----|-----------------------------------|---------------|
| 7  | Апробація результатів дослідження | 2.12.2019 р.  |
| 8  | Публікації                        | 06.12.2019 р. |
| 9  | Попередній захист                 | 07.12.2019 р. |
| 10 | Захист МКР                        | 13.12.2019 р. |

#### 7.2 Графічні матеріали:

|                            |                      |
|----------------------------|----------------------|
| – схема взаємодії програм: | «02 » грудня 2019 р. |
| – схема даних:             | «04 » грудня 2019 р. |
| – схеми алгоритмів:        | «05 » грудня 2019 р. |
| – плакати демонстраційні:  | «05 » грудня 2019 р. |

### 8. Порядок контролю і приймання.

8.1. Хід виконання магістерської кваліфікаційної роботи контролюється керівником роботи, консультантами з економічної частини. Рубіжний контроль провести до «06» грудня 2019 р

8.2. Атестація проекту здійснюється на попередньому захисті. Попередній захист магістерської кваліфікаційної роботи провести до «07» грудня 2019 р.

8.3. Підсумкове рішення щодо оцінки якості виконання магістерської кваліфікаційної роботи приймається на засіданні ДЕК. Захист магістерської кваліфікаційної роботи провести «13» грудня 2019 р.



## Додаток Б (обов'язковий)

### Лістинги основних модулів програми

#### Лістинг модуля “TextAnalyze.cpp”

```

#include "stdafx.h"
#include "TextAnalyze.h"
#include "TextAnalyzeDlg.h"

#ifdef _DEBUG
#define new DEBUG_NEW
#undef THIS_FILE
static char THIS_FILE[] = __FILE__;
#endif

////////////////////////////////////
// CTextAnalyzeApp

BEGIN_MESSAGE_MAP(CTextAnalyzeApp, CWinApp)
//{{AFX_MSG_MAP(CTextAnalyzeApp)
// NOTE - the ClassWizard will add and remove mapping macros here.
// DO NOT EDIT what you see in these blocks of generated code!
//}}AFX_MSG
ON_COMMAND(ID_HELP, CWinApp::OnHelp)
END_MESSAGE_MAP()

////////////////////////////////////
// CTextAnalyzeApp construction

CTextAnalyzeApp::CTextAnalyzeApp()
{
    // TODO: add construction code here,
    // Place all significant initialization in InitInstance
}

////////////////////////////////////
// The one and only CTextAnalyzeApp object

CTextAnalyzeApp theApp;

////////////////////////////////////
// CTextAnalyzeApp initialization

BOOL CTextAnalyzeApp::InitInstance()
{
    // Standard initialization
    // If you are not using these features and wish to reduce the size
    // of your final executable, you should remove from the following
    // the specific initialization routines you do not need.

#ifdef _AFXDLL
    Enable3dControls();           // Call this when using MFC in a shared DLL
#else
    Enable3dControlsStatic();     // Call this when linking to MFC statically
#endif

    CTextAnalyzeDlg dlg;
    m_pMainWnd = &dlg;
    int nResponse = dlg.DoModal();
    if (nResponse == IDOK)
    {
        // TODO: Place code here to handle when the dialog is
        // dismissed with OK
    }
    else if (nResponse == IDCANCEL)
    {
        // TODO: Place code here to handle when the dialog is
        // dismissed with Cancel
    }

    // Since the dialog has been closed, return FALSE so that we exit the
    // application, rather than start the application's message pump.

```

```

    return FALSE;
}

```

## Лістинг модуля “TextAnalyzeDlg.cpp”

```

#include "stdafx.h"
#include "TextAnalyze.h"
#include "TextAnalyzeDlg.h"
#include "DiagramDlg.h"

#ifdef _DEBUG
#define new DEBUG_NEW
#undef THIS_FILE
static char THIS_FILE[] = __FILE__;
#endif

#define BUFF_SIZE 1024
#define FILTER "Text files (*.txt)|*.txt|All files|*.*" // file filter

////////////////////////////////////
// CAboutDlg dialog used for App About

class CAboutDlg : public CDialog
{
public:
    CAboutDlg();

// Dialog Data
   //{{AFX_DATA(CAboutDlg)
    enum { IDD = IDD_ABOUTBOX };
    //}}AFX_DATA

    // ClassWizard generated virtual function overrides
   //{{AFX_VIRTUAL(CAboutDlg)
protected:
    virtual void DoDataExchange(CDataExchange* pDX); // DDX/DDV support
    //}}AFX_VIRTUAL

// Implementation
protected:
   //{{AFX_MSG(CAboutDlg)
    //}}AFX_MSG
    DECLARE_MESSAGE_MAP()
};

CAboutDlg::CAboutDlg() : CDialog(CAboutDlg::IDD)
{
   //{{AFX_DATA_INIT(CAboutDlg)
    //}}AFX_DATA_INIT
}

void CAboutDlg::DoDataExchange(CDataExchange* pDX)
{
    CDialog::DoDataExchange(pDX);
   //{{AFX_DATA_MAP(CAboutDlg)
    //}}AFX_DATA_MAP
}

BEGIN_MESSAGE_MAP(CAboutDlg, CDialog)
   //{{AFX_MSG_MAP(CAboutDlg)
    // No message handlers
    //}}AFX_MSG_MAP
END_MESSAGE_MAP()

////////////////////////////////////
// CTextAnalyzeDlg dialog

CTextAnalyzeDlg::CTextAnalyzeDlg(CWnd* pParent /*=NULL*/)
: CDialog(CTextAnalyzeDlg::IDD, pParent)
{
   //{{AFX_DATA_INIT(CTextAnalyzeDlg)
    m_text1 = _T("");
    m_text2 = _T("");
    //}}AFX_DATA_INIT
    // Note that LoadIcon does not require a subsequent DestroyIcon in Win32
    m_hIcon = AfxGetApp()->LoadIcon(IDR_MAINFRAME);
}

void CTextAnalyzeDlg::DoDataExchange(CDataExchange* pDX)

```

```

{
    CDialog::DoDataExchange(pDX);
    //{{AFX_DATA_MAP(CTextAnalyzeDlg)
    DDX_Text(pDX, IDC_EDIT1, m_text1);
    DDX_Text(pDX, IDC_EDIT2, m_text2);
    //}}AFX_DATA_MAP
}

BEGIN_MESSAGE_MAP(CTextAnalyzeDlg, CDialog)
    //{{AFX_MSG_MAP(CTextAnalyzeDlg)
    ON_WM_SYSCOMMAND()
    ON_WM_PAINT()
    ON_WM_QUERYDRAGICON()
    ON_COMMAND(IDM_FILE_EXIT, OnFileExit)
    ON_COMMAND(IDM_TEXT_ANALYZE, OnTextAnalyze)
    ON_COMMAND(ID_FILE_OPENTEXT1, OnFileOpenText1)
    ON_COMMAND(ID_FILE_OPENTEXT2, OnFileOpenText2)
    ON_COMMAND(ID_HELP_ABOUT, OnHelpAbout)
    ON_COMMAND(ID_TEXT_CLEARTEXT1, OnTextClearText1)
    ON_COMMAND(ID_TEXT_CLEARTEXT2, OnTextClearText2)
    ON_WM_SIZE()
    ON_COMMAND(ID_TEXT_NORMALIZETEXT1, OnTextNormalizeText1)
    ON_COMMAND(ID_TEXT_NORMALIZETEXT2, OnTextNormalizeText2)
    //}}AFX_MSG_MAP
END_MESSAGE_MAP()

////////////////////////////////////
// CTextAnalyzeDlg message handlers

BOOL CTextAnalyzeDlg::OnInitDialog()
{
    CDialog::OnInitDialog();

    // Add "About..." menu item to system menu.

    // IDM_ABOUTBOX must be in the system command range.
    ASSERT((IDM_ABOUTBOX & 0xFFF0) == IDM_ABOUTBOX);
    ASSERT(IDM_ABOUTBOX < 0xF000);

    CMenu* pSysMenu = GetSystemMenu(FALSE);
    if (pSysMenu != NULL)
    {
        CString strAboutMenu;
        strAboutMenu.LoadString(IDS_ABOUTBOX);
        if (!strAboutMenu.IsEmpty())
        {
            pSysMenu->AppendMenu(MF_SEPARATOR);
            pSysMenu->AppendMenu(MF_STRING, IDM_ABOUTBOX, strAboutMenu);
        }
    }

    // Set the icon for this dialog. The framework does this automatically
    // when the application's main window is not a dialog
    SetIcon(m_hIcon, TRUE); // Set big icon
    //SetIcon(m_hIcon, FALSE); // Set small icon

    return TRUE; // return TRUE unless you set the focus to a control
}

void CTextAnalyzeDlg::OnSysCommand(UINT nID, LPARAM lParam)
{
    if ((nID & 0xFFF0) == IDM_ABOUTBOX)
    {
        CAboutDlg dlgAbout;
        dlgAbout.DoModal();
    }
    else
        CDialog::OnSysCommand(nID, lParam);
}

// If you add a minimize button to your dialog, you will need the code below
// to draw the icon. For MFC applications using the document/view model,
// this is automatically done for you by the framework.

void CTextAnalyzeDlg::OnPaint()
{
    if (IsIconic())
    {
        CPaintDC dc(this); // device context for painting

        SendMessage(WM_ICONERASEBKGND, (LPARAM) dc.GetSafeHdc(), 0);
    }
}

```

```

        // Center icon in client rectangle
        int cxIcon = GetSystemMetrics(SM_CXICON);
        int cyIcon = GetSystemMetrics(SM_CYICON);
        CRect rect;
        GetClientRect(&rect);
        int x = (rect.Width() - cxIcon + 1) / 2;
        int y = (rect.Height() - cyIcon + 1) / 2;

        // Draw the icon
        dc.DrawIcon(x, y, m_hIcon);
    }
    else
        CDialog::OnPaint();
}

// The system calls this to obtain the cursor to display while the user drags
// the minimized window.
HCURSOR CTextAnalyzeDlg::OnQueryDragIcon()
{
    return (HCURSOR) m_hIcon;
}

void CTextAnalyzeDlg::OnFileExit()
{
    OnOK();
}

void CTextAnalyzeDlg::OnHelpAbout()
{
    CAboutDlg dlgAbout;
    dlgAbout.DoModal();
}

void CTextAnalyzeDlg::OnFileOpenText1()
{
    OpenFile(m_text1, "Open Text 1");
}

void CTextAnalyzeDlg::OnFileOpenText2()
{
    OpenFile(m_text2, "Open Text 2");
}

bool CTextAnalyzeDlg::OpenFile(CString &s, const char *szTitle)
{
    char szFile[MAX_PATH]={0};
    OPENFILENAME ofn;
    memset(&ofn, 0, sizeof(OPENFILENAME));
    ofn.hwndOwner = m_hWnd;
    ofn.lStructSize = sizeof(OPENFILENAME);
    ofn.Flags = OFN_HIDEREADONLY | OFN_EXPLORER | OFN_FILEMUSTEXIST | OFN_PATHMUSTEXIST;
    ofn.lpstrFilter = FILTER;
    ofn.nFilterIndex= 1;
    ofn.lpstrFile = szFile;
    ofn.nMaxFile = MAX_PATH;
    ofn.lpstrTitle = szTitle;

    if( GetOpenFileName(&ofn) )
    {
        UpdateData(1);
        CFile f;
        char szBuff[BUFF_SIZE+1];
        int read;

        s.Empty();

        if( f.Open(szFile, CFile::modeRead) )
        {
            do
            {
                read = f.Read(szBuff, BUFF_SIZE);
                szBuff[read] = 0;

                s += szBuff;
            } while( read );

            f.Close();
            UpdateData(0);

            return true;
        }
    }
}

```

```

        return false;
    }

void CTextAnalyzeDlg::OnTextClearText1()
{
    UpdateData(1);
    m_text1.Empty();
    UpdateData(0);
}

void CTextAnalyzeDlg::OnTextClearText2()
{
    UpdateData(1);
    m_text2.Empty();
    UpdateData(0);
}

void CTextAnalyzeDlg::OnTextNormalizeText1()
{
    CText text1;

    BeginWaitCursor();

    UpdateData(1);

    text1.OpenText(m_text1);
    m_text1 = text1.GetText();

    UpdateData(0);

    EndWaitCursor()
}

void CTextAnalyzeDlg::OnTextNormalizeText2()
{
    CText text2;

    BeginWaitCursor();

    UpdateData(1);

    text2.OpenText(m_text2);
    m_text2 = text2.GetText();

    UpdateData(0);

    EndWaitCursor();
}

void CTextAnalyzeDlg::OnSize(UINT nType, int cx, int cy)
{
    CDialog::OnSize(nType, cx, cy);

    static bool b = false;

    if( cx && cy && b )
    {
        CRect r;
        GetClientRect(&r);

        r.DeflateRect(10, 10, 10, 10);
        int w = r.Width(), h = r.Height()/2 - 5;

        GetDlgItem(IDC_EDIT1)->MoveWindow(r.left, r.top, w, h);
        GetDlgItem(IDC_EDIT2)->MoveWindow(r.left, r.top + h + 10, w, h);
    }
    b = true;
}

void CTextAnalyzeDlg::OnTextAnalyze()
{
    CText text1, text2;

    BeginWaitCursor();

    UpdateData(1);

    text1.OpenText(m_text1);
    text2.OpenText(m_text2);

    int nSentence = text1.GetCount();
    int nWords = text1.GetWordsCount();
}

```

```

if( nSentence )
{
    CString sInfo, sTmp;
    STAT_INFO stat;
    CDiagramDlg dlg;
    float x[4], y[4];

    sInfo.Format("Всього слів = %d\nКількість речень = %d\n", nWords, nSentence);
    sInfo += "-----\n";

    for(int w=1; w <= 4; ++w)
    {
        stat = GetStat(&text1, &text2, w);

        sTmp.Format("Для %d-грам:\n", w);
        sInfo += sTmp;

        sTmp.Format("Співпало = %d\nВсього = %d\nТочність = %.4f\n", stat.nMatch, stat.nAllCount,
stat.AverageMatchPercent);
        sInfo += sTmp;

        x[w-1] = stat.AverageMatchPercent;
        y[w-1] = w;
    }

    EndWaitCursor();

    MessageBox(sInfo);

    dlg.m_x = x;
    dlg.DoModal();
}
else EndWaitCursor();
}

bool IsEqual(CString s1, CString s2)
{
    s1.MakeLower();
    s2.MakeLower();

    return s1 == s2;
}

STAT_INFO CTextAnalyzeDlg::GetStat(CText *ptext1, CText *ptext2, int nWords)
{
    STAT_INFO stat;
    memset(&stat, 0, sizeof(stat));

    int count = min(ptext1->GetCount(), ptext2->GetCount());
    int wordCount1, wordCount2;
    bool b;
    CSentence *psen1, *psen2;

    ptext1->SetMarked(false);
    ptext2->SetMarked(false);

    for(int i=0; i < count; ++i)
    {
        psen1 = ptext1->Get(i);
        psen2 = ptext2->Get(i);

        wordCount1 = psen1->GetCount();
        wordCount2 = psen2->GetCount();

        if( wordCount1 - nWords + 1 <= 0 )
            continue;

        stat.nAllCount += wordCount1 - nWords + 1;

        for (int j=0; j < wordCount1-nWords+1; ++j)
        {
            for (int k=0; k < wordCount2-nWords+1; ++k)
            {
                b = true;

                for(int w=0; w < nWords; ++w)
                    //if( psen1->Get(j+w) != psen2->Get(k+w) )
                    //if( !IsEqual( psen1->Get(j+w), psen2->Get(k+w) ) )
                    //if( psen1->GetWordPtr(j+w)->IsEqual( *psen2->GetWordPtr(k+w) ) )
                    if( !((*psen1->GetWordPtr(j+w)) == (*psen2->GetWordPtr(k+w))) )

```

```

        {
            b = false;
            break;
        }
        else
            if(nWords > 1) psen2->SetMarked(false);

        if( b )
        {
            ++stat.nMatch;
            break;
        }
    }
}

if( stat.nAllCount )
    stat.AverageMatchPercent = (float)stat.nMatch / stat.nAllCount;

return stat;
}

////////////////////////////////////< CText >////////////////////////////////////

// open and parse text
void CText::OpenText(const CString &sText)
{
    m_text.clear();

    CSentence sen;
    CString sWord;
    char *s = (char*)(const char*)sText;

    while( *s )
    {
        switch( *s )
        {
            // word separators
            case '!':
            case ';':
            case '\n':
            case '\r':
            case '\t':
                if( !sWord.IsEmpty() )
                {
                    sen.AddWord(sWord);
                    sWord.Empty();
                }
                break;

            // sentence separators
            case '!':
            case '!':
            case '?':
                if( !sWord.IsEmpty() )
                {
                    sen.AddWord(sWord);
                    sWord.Empty();
                }
                if( sen.GetCount() )
                {
                    m_text.push_back(sen);
                    sen.Clear();
                }
                break;

            // word
            default:
                sWord += *s;
                break;
        }
        ++s;
    }
    if( !sWord.IsEmpty() )
        sen.AddWord(sWord);

    if( sen.GetCount() )
        m_text.push_back(sen);
}

CString CText::GetText()
{

```

```

    CString sText;

    int SenCount = m_text.size();
    for (int i=0; i < SenCount; ++i)
    {
        sText += m_text[i].GetSentence();

        if( i < SenCount-1 )
            sText += ' ';
    }
    return sText;
}

int CText::GetWordsCount()
{
    int words = 0;

    for(int i=0; i < m_text.size(); ++i)
        words += m_text[i].GetCount();

    return words;
}

void CText::SetMarked(bool bm)
{
    for(int i=0; i < m_text.size(); ++i)
        m_text[i].SetMarked(bm);
}

/////////////////////////////////< CSentence >////////////////////////////////

CString CSentence::GetSentence()
{
    CString sSen;

    int WordCount = m_Words.size();
    for (int i=0; i < WordCount; ++i)
    {
        sSen += m_Words[i];

        if( i < WordCount-1 )
            sSen += ' ';
    }
    sSen += ' ';
    return sSen;
}

```



Додаток В  
(Обов'язковий)

**ЗАТВЕРДЖУЮ**  
Завідувач кафедри КСУ  
д.т.н., проф. В.М. Дубовой

« \_\_\_\_\_ » \_\_\_\_\_ 2019 р.

ПЕРЕЛІК  
ГРАФІЧНИХ МАТЕРІАЛІВ

для захисту магістерської кваліфікаційної роботи  
на тему

**РОЗРОБКА МЕТОДІВ ОЦІНКИ ЕФЕКТИВНОСТІ АВТОМАТИЧНОГО  
ПЕРЕКЛАДУ МОВИ В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ АВТОМАТИЗАЦІЇ**

1. Узагальнена схема оцінювання систем МП.
2. Математична модель оцінювання машинного перекладу. ММЕМ - Модифікований метод N-грам із використанням морфем
3. Математична модель оцінювання машинного перекладу. ISKM - метод оцінки ефективності систем МП на основі інформаційно-статистичного критерію і моделі потенціальної системи
4. Схема алгоритму приведення тексту до зручного для оцінювання вигляду.
5. Схема алгоритму порівняння двох текстів.
6. Програмне забезпечення системи оцінювання МП. Схема роботи програми-аналізатора.
7. Програмне забезпечення системи оцінювання МП. Файлова структура програми TextAnAlyze.
8. Результати експериментальних досліджень.
9. Оцінка адекватності методів і моделей оцінювання якості МП.

Розробив: Власюк М.В. \_\_\_\_\_

(підпис) (дата)

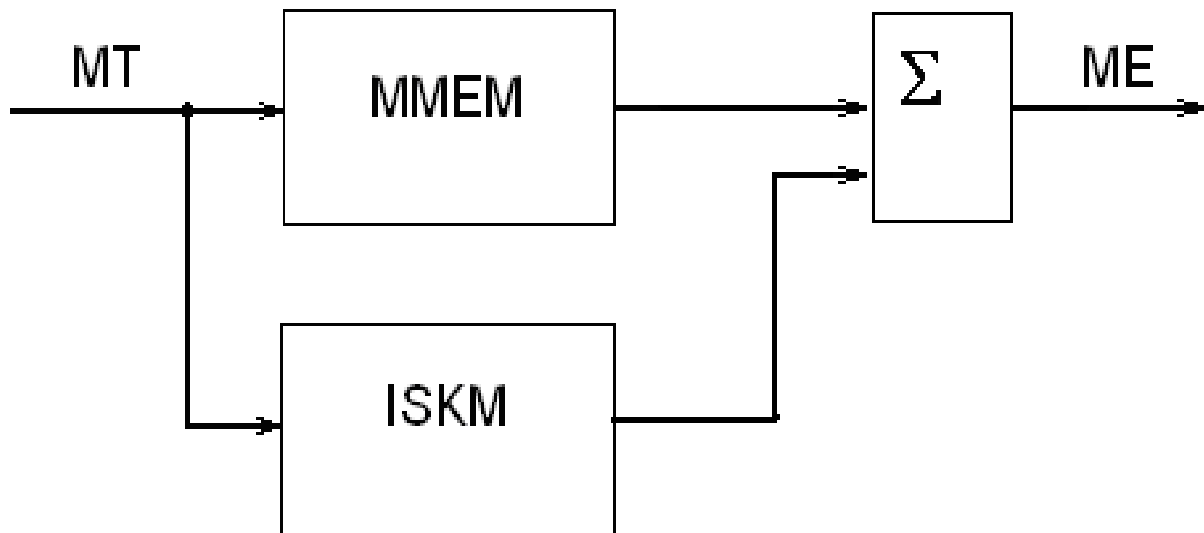
Перевірив: Биков М. М. \_\_\_\_\_

(підпис) (дата)

Рецензент: Севастьянов В. М. \_\_\_\_\_

(підпис) (дата)

## УЗАГАЛЬНЕНА СХЕМА ОЦІНЮВАННЯ СИСТЕМ МП



*MT* – машинний переклад;

*ME* – автоматична оцінка машинного перекладу;

*MMEM* - модифікований метод N-грам із зміненою точністю та використанням морфем;

*ISKM* - метод оцінки ефективності систем МП на основі інформаційно-статистичного критерію і моделі потенціальної системи.

# МАТЕМАТИЧНА МОДЕЛЬ ОЦІНЮВАННЯ МАШИННОГО ПЕРЕКЛАДУ

## ММЕМ - Модифікований метод N-грам із використанням морфем

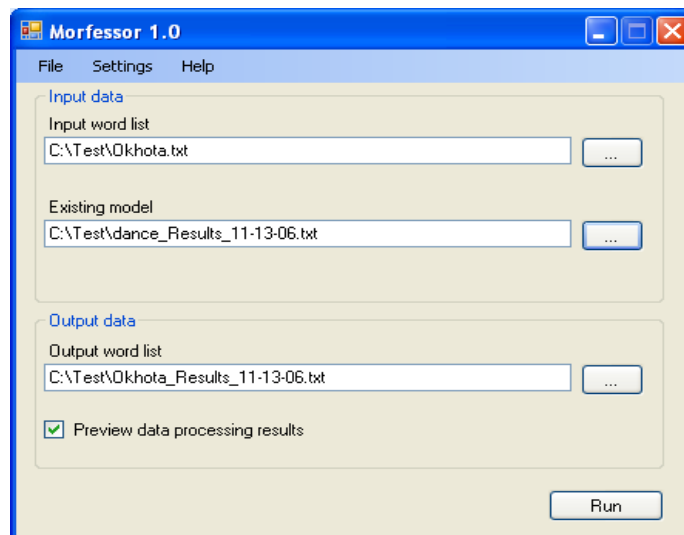
$$p_n = \frac{\sum_{C \in (\text{Candidates})} \sum_{n\text{-gram} \in C} \text{Count}_{clip}(n\text{-gram})}{\sum_{C' \in (\text{Candidates})} \sum_{n\text{-gram}' \in C'} \text{Count}_{clip}(n\text{-gram}' )}$$

$\sum_{C \in (\text{Candidates})} \sum_{n\text{-gram} \in C} \text{Count}_{clip}(n\text{-gram})$  - кількість N-грам машинного перекладу, що співпали з N-грамами еталонного перекладу, в якості N-грам використовуються морфеми;

$\sum_{C' \in (\text{Candidates})} \sum_{n\text{-gram}' \in C'} \text{Count}_{clip}(n\text{-gram}' )$  - загальна кількість N-грам (морфем) еталонного перекладу.

**Недолік методу** - використовує лише статистичні дані послівно, і не враховує синтаксичну інформацію про мову. Використання тільки послівного аналізу для їх оцінювання не дозволяє отримати задовільних результатів, тому виникає необхідність використання додаткової інформації. В якості якої може слугувати лексична і лінгвістична інформація. Лексичну інформацію зручно використовувати тоді, коли елементами розпізнавання будуть морфи – мінімальні змістовно розділимі частини слова (наприклад, префікс, корінь, суфікс, закінчення).

## Графічний інтерфейс програми Morfessor – 1.0



# МАТЕМАТИЧНА МОДЕЛЬ ОЦІНЮВАННЯ МАШИННОГО ПЕРЕКЛАДУ

**ISKM - метод оцінки ефективності систем МП на основі інформаційно-статистичного критерію і моделі потенціальної системи**

$E = \frac{I_{MT}}{I_P}$  - модифікований до систем штучного інтелекту інформаційний критерій,

де  $I_{MT}$  - інформативність перекладеного машинним перекладачем тексту,

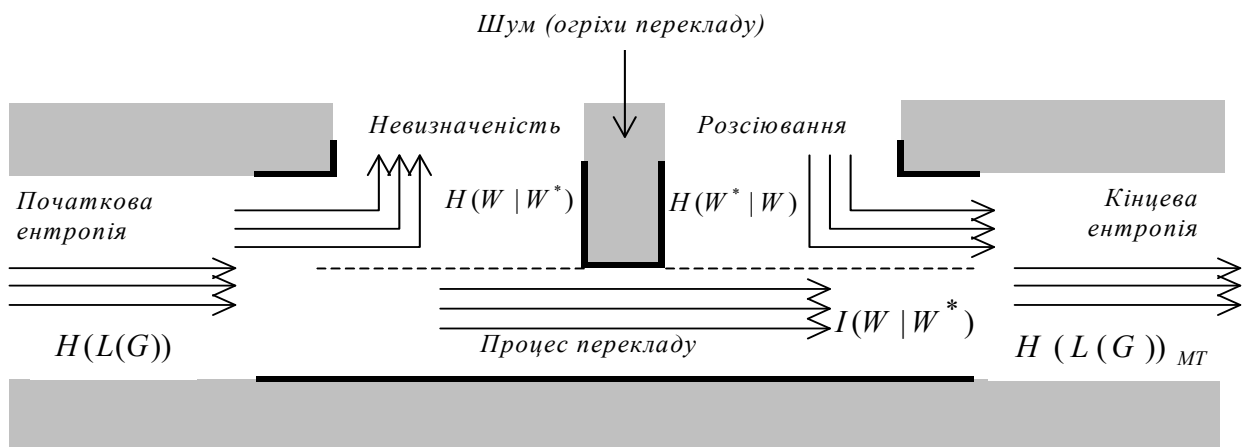
$$I_{MT} = H(L(G)) - H(L(G))_{MT},$$

$$I_P = H(L(G)) - H(L(G))_P,$$

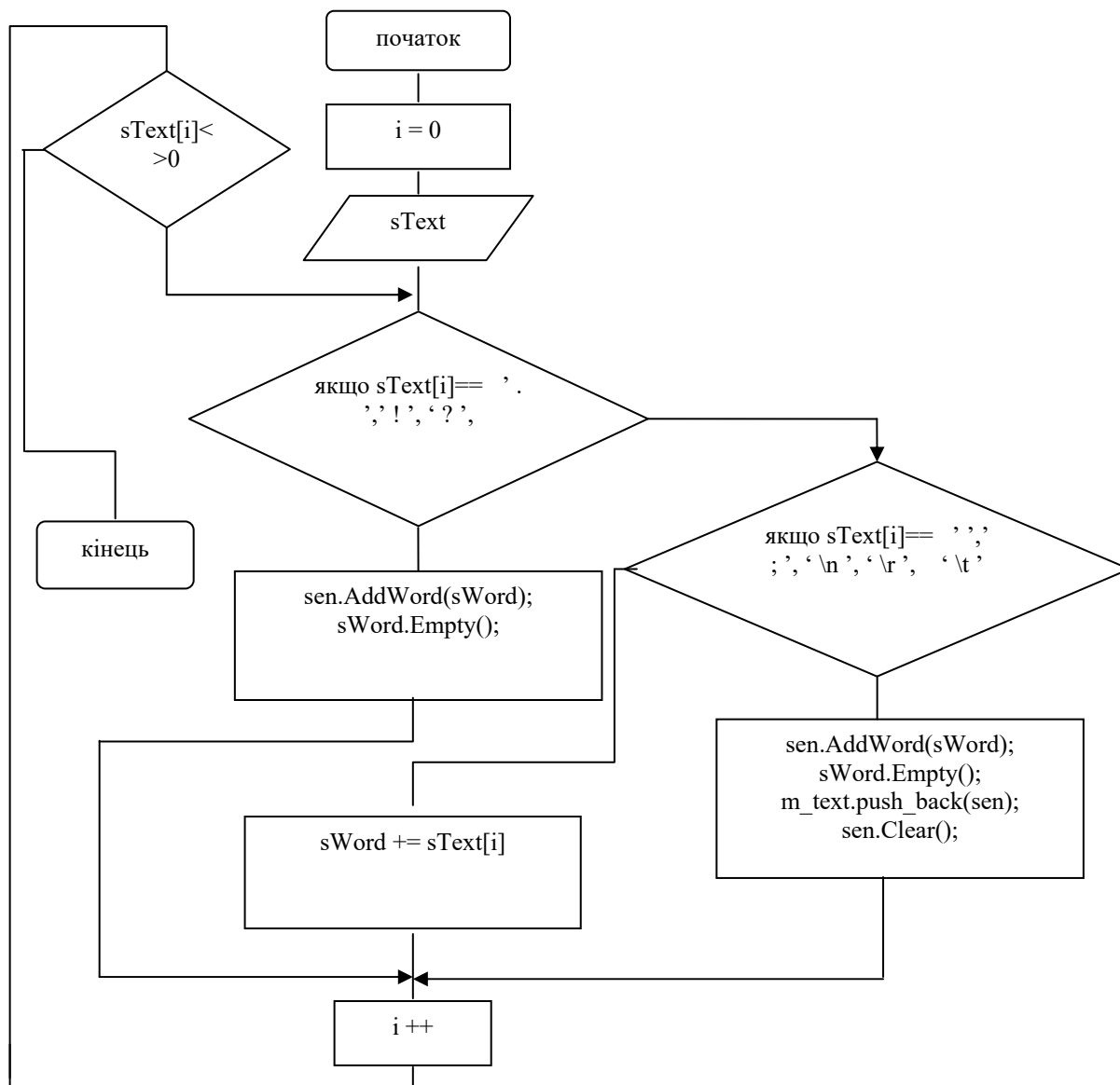
де  $H(L(G))$  - ентропія мови;  $H(L(G))_{MT}$  - ентропія тексту, перекладеного машинним перекладачем;  $H(L(G))_P$  - ентропія тексту, перекладеного потенціальною (ідеальною) системою перекладу. В якості потенціальної системи перекладу вибрано людину – професіонального перекладача, тому в першому наближенні можна вважати, що

$$H(L(G))_P = H(L(G)).$$

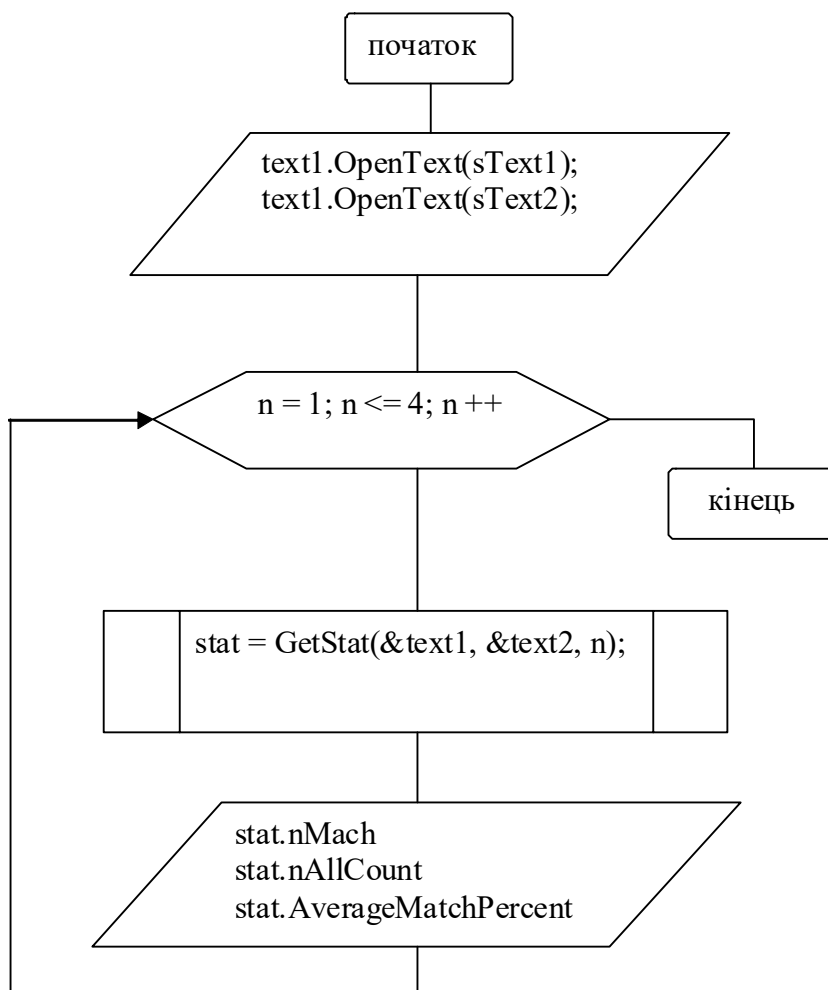
**Представлення процесу машинного перекладу тексту каналом з шумами**



## СХЕМА АЛГОРИТМУ ПРИВЕДЕННЯ ТЕКСТУ ДО ЗРУЧНОГО ДЛЯ ОЦІНЮВАННЯ ВИГЛЯДУ

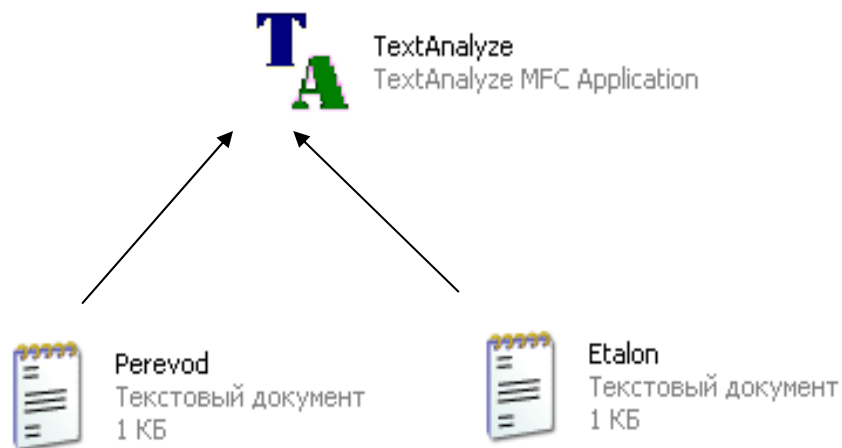


## СХЕМА АЛГОРИТМУ ПОРІВНЯННЯ ДВОХ ТЕКСТІВ



# ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ СИСТЕМИ ОЦІНЮВАННЯ МП

## Схема роботи програми-аналізатора

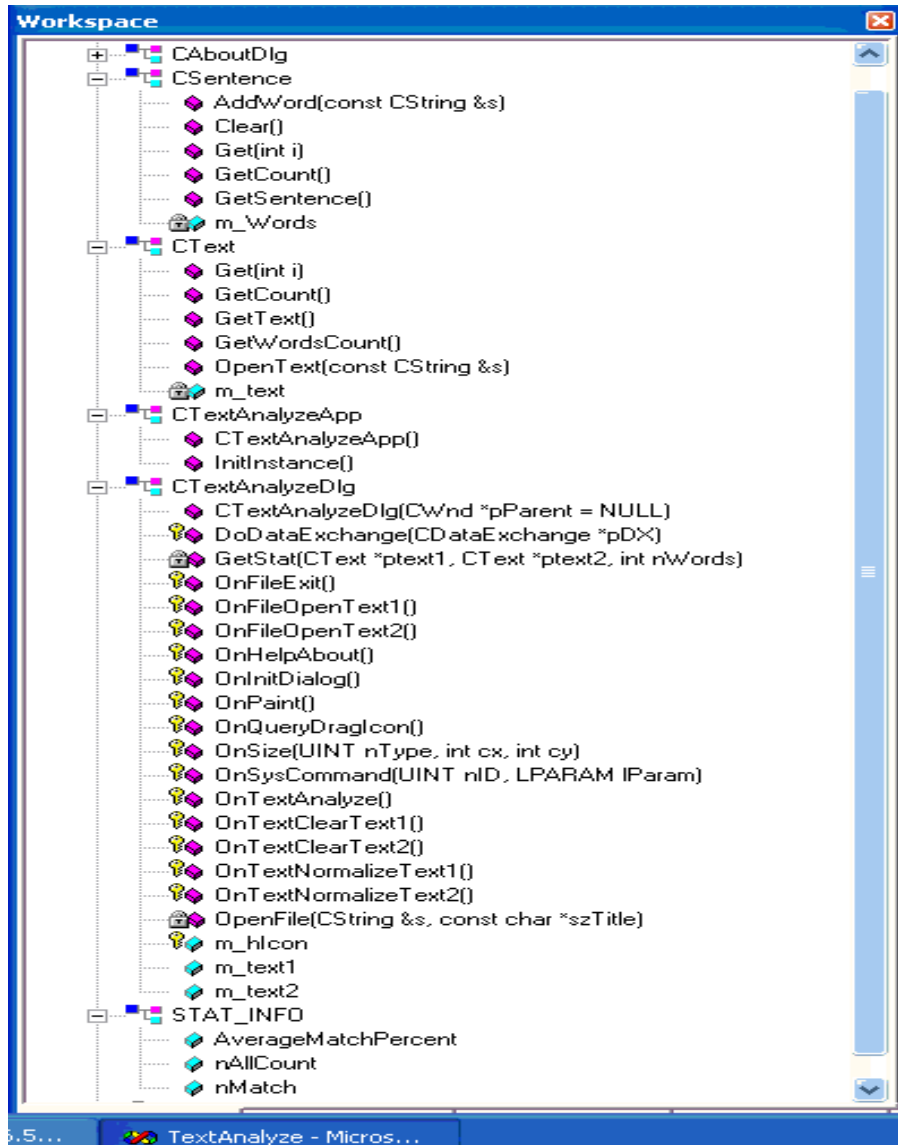


Розроблене програмне забезпечення складається з модулів:

- TextAnalyze  
e.exe – головна програма;
- Perevod –  
модуль, в якому зберігається машинний переклад.
- Etalon –  
модуль, в якому зберігається переклад експерта.

# ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ СИСТЕМИ ОЦІНЮВАННЯ МП

## Файлова структура програми TextAnalyze



Опис класів головної програми.

class CSentence – клас обробки речення в програмі.

class CText – клас обробки тексту в програмі.

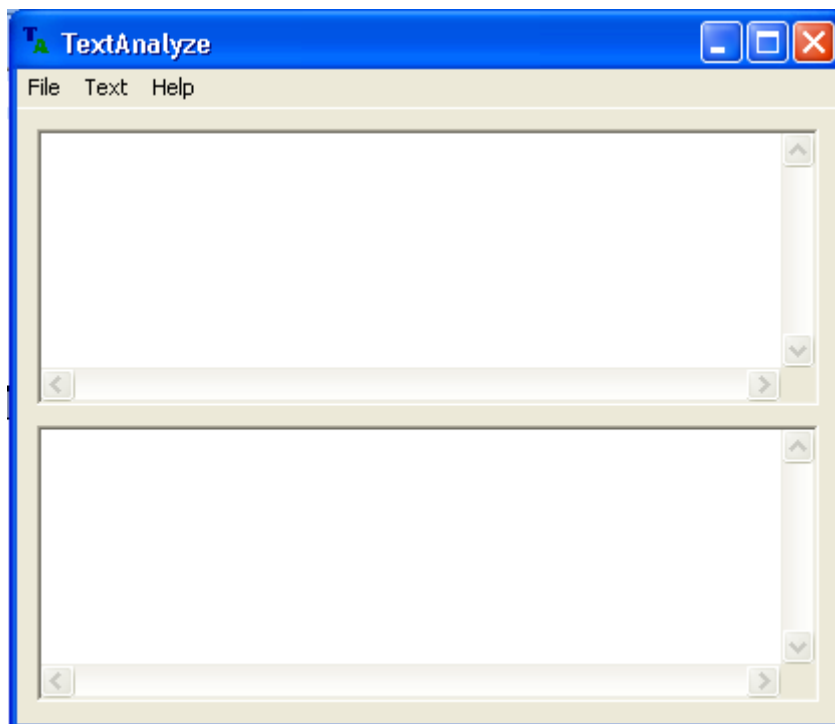
class CTextAnalyzeApp – клас для створення об'єкту CText і передачі керування.

class CTextAnalyzeDlg – клас проведення аналізу співпадінь у двох запропонованих текстах.



## РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТАЛЬНИХ ДОСЛІДЖЕНЬ

### Графічний інтерфейс програми



### Еталонний переклад тексту “Wallaby” людиною

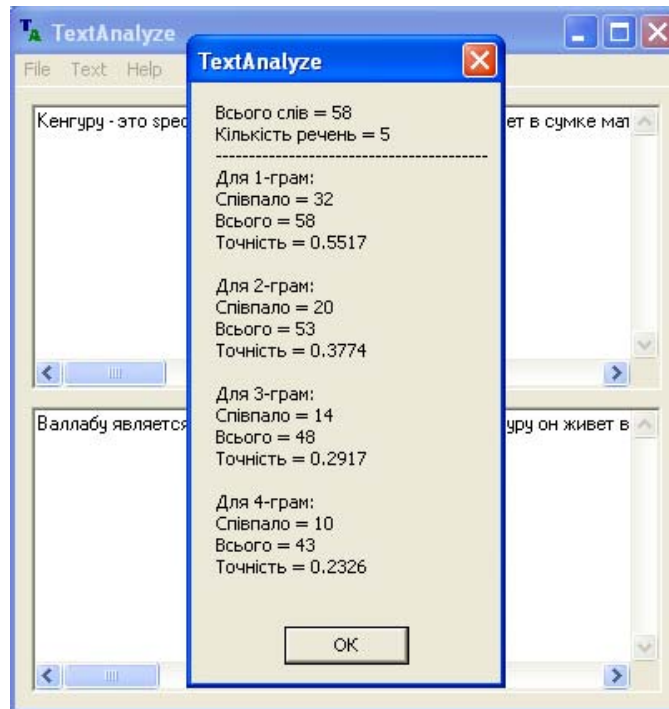
Валлабу являється різновидністю кенгуру. Як і кенгуру він живе в сумці матері перші кілька місяців після народження до тих пор поки він зможе стояти на власних ногах. Це допомагає йому вижити в важких умовах. Валлабу можуть жити 8 або 9 місяців без свіжої води. Воду, яку їм потрібно, вони отримують з рослин, тому їм не потрібно бути поруч з водою.

### Переклад тексту програмою Pragma

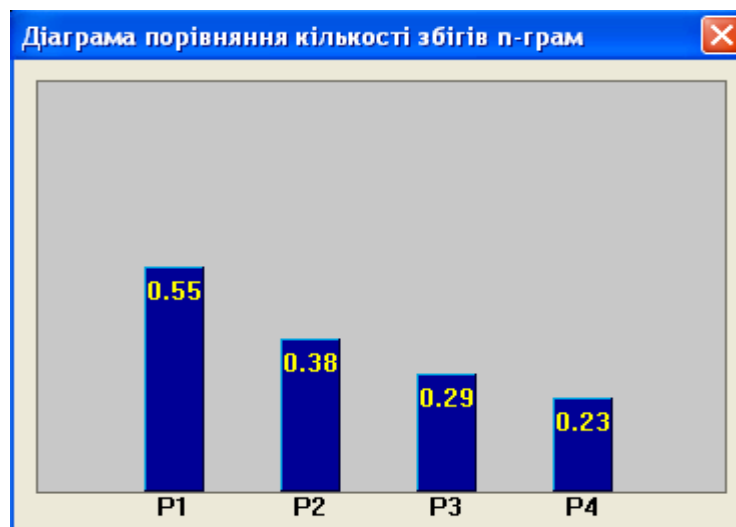
“Кенгуру - це species кенгуру. Подібно кенгуру це живе в сумці матері перші кілька місяців після початку, поки це може постояти на власних ногах. Це допомогає їм, щоб витримати важкі умови. Кенгуру можуть жити впродовж 8 або 9 місяців без прісної води. Вони отримують воду їм потрібно від заводів так що їм не потрібно бути поруч з водою.”

## Результат оцінювання перекладу програмою Pragma:

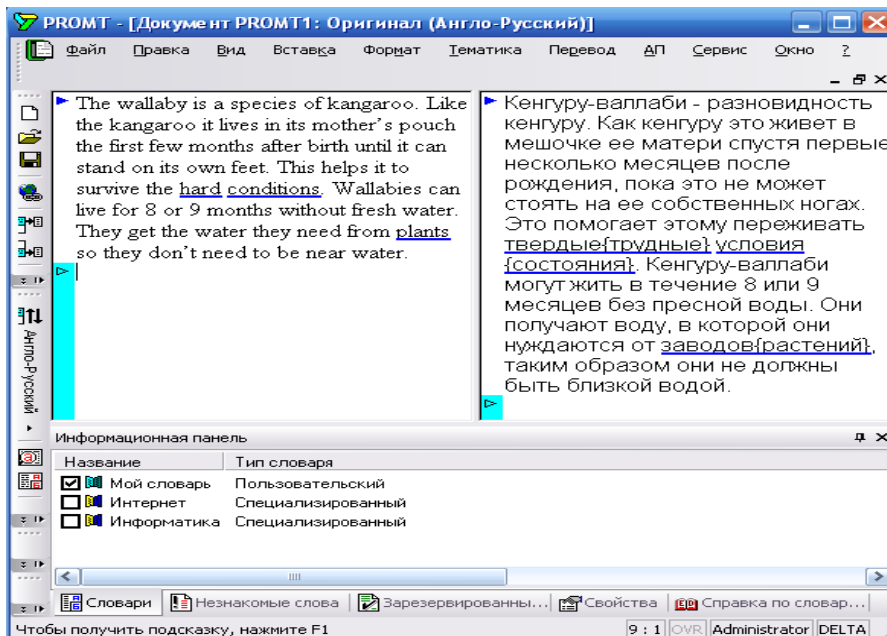
Завантажуємо Файл → OpenText1 → Pragma.txt та Файл → OpenText2 → Exp.txt, після натискаємо Text → Analyze.



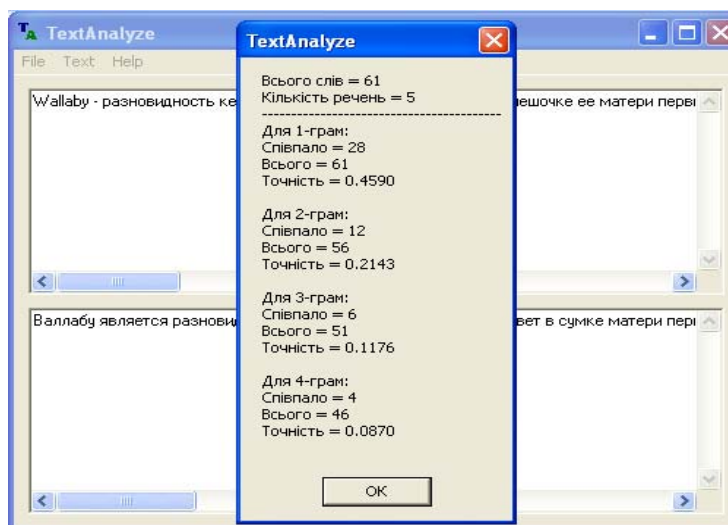
Діаграма порівняння кількості збігів N-грам для системи Pragma



## Переклад тексту програмою Promt



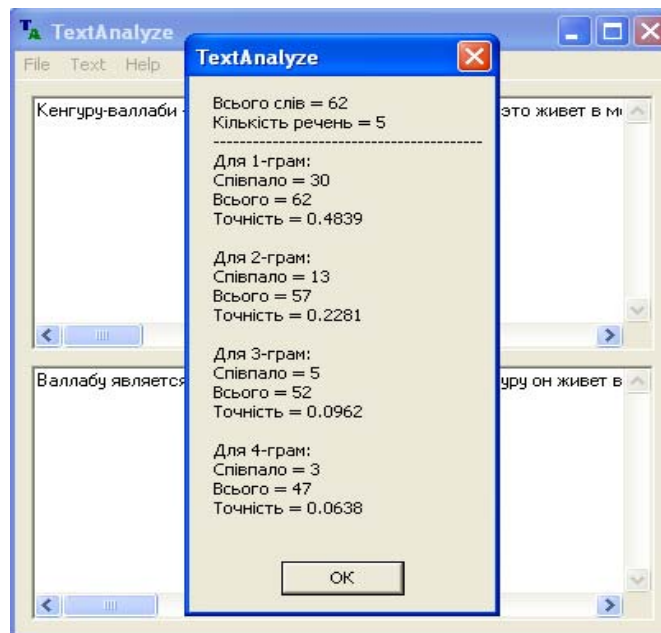
## Результати оцінювання перекладу програмою Promt



## Діаграма порівняння кількості збігів N-грам для системи Pragma



## Результат оцінювання перекладу програмою Google Translate:



## Діаграма порівняння кількості збігів N-грам для програми Google Translate



# ОЦІНКА АДЕКВАТНОСТІ МЕТОДІВ І МОДЕЛЕЙ ОЦІНЮВАННЯ ЯКОСТІ МП

## Алгоритм дослідження адекватності

1. Одержання результатів оцінки систем МП за допомогою розробленого методу та за допомогою експерта.
2. Знаходження кореляційної залежності між оцінкою якості перекладу яку дав експерт та оцінкою отриманою за розробленим алгоритмом.
3. Оцінити щільність зв'язку між відповідними ознаками за кореляційним відношенням, обчислити коефіцієнт детермінації.
4. Зробити висновок про адекватність оцінки за розробленим методом, оцінці перекладу експертом.

## Результати оцінювання систем МП за допомогою розробленого алгоритму (МЕ) та за допомогою експерта(НЕ)

| Система МП       | МЕ       | НЕ    |
|------------------|----------|-------|
| pragma           | 0.351481 | 3.977 |
| Google Translate | 0.344167 | 3.94  |
| Prompt           | 0.256204 | 3.858 |

## Кореляційна залежність між оцінкою якості переклад, яку дав експерт (НЕ), та оцінкою отриманою за розробленим алгоритмом (МЕ)

Обчислений коефіцієнт кореляції між оцінками:  $r = 0,95554$ .

Згідно критерію Івахненко ( $r > 0,73$ ), розроблені методи і моделі оцінювання систем МП є адекватними