

УДК 004.83

DOI: 10.18523/2617-3808.2022.5.54-61

Крамов А. А., Погорілий С. Д.

## ЗАСТОСУВАННЯ МЕТОДУ ВИЯВЛЕННЯ ЕЛЕМЕНТІВ ДИСФЛЮЕНТНОСТІ МОВЛЕННЯ ДЛЯ МАШИННОГО ПЕРЕКЛАДУ ТРАНСКРИБЦІЙ УСНОЇ МОВИ

У роботі розглянуто проблему здійснення нейронного машинного перекладу результатів транскрибування усної мови, які містять елементи дисфлюентності мовлення, з англійської мови на українську. Проаналізовано наявні методи та програмні засоби виявлення елементів дисфлюентності мовлення в англійських текстах. Створено синтетичний розмічений корпус, що містить оригінальні версії документів та їхні модифіковані версії відповідно до визначених типів дисфлюентності мовлення. Здійснено експериментальну перевірку ефективності застосування методу виявлення елементів дисфлюентності мовлення для вдосконалення перекладу усної мови на основі застосування моделей нейронного машинного перекладу для пари англійської та української мов. Отримані результати свідчать про доцільність використання зазначеного методу для попереднього оброблення результатів транскрибування діалогів для створення коректних когерентних перекладів за допомогою зазначених моделей машинного перекладу.

**Ключові слова:** оброблення природної мови, нейронний машинний переклад, дисфлюентність мовлення, оброблення усної мови.

### Вступ

Виконання машинного перекладу – автоматизованого перекладу текстів з однієї природної мови на іншу – належить до основних задач галузі оброблення природної мови (англ. NLP – natural language processing), що підтверджується наявністю значної кількості відповідних академічних робіт [1; 3], а також комерційних рішень [24] (Google Translate, DeepL Translate тощо). Крім того, актуальним напрямом є дослідження створення моделей багатомовного перекладу з можливістю оброблення мов з низькою кількістю (чи навіть без) відповідних розмічених корпусів, про що свідчить актуальна робота компанії Meta AI [18]. Розвиток парадигми глибокого навчання та отримані результати точності застосування багатопарових нейронних мереж різної архітектури зумовили поступовий перехід від моделей *статистичного* (англ. SMT – statistical machine translation) до моделей *нейронного* (англ.

NMT – neural machine translation) машинного перекладу. Моделі статистичного машинного перекладу ґрунтуються на використанні окремих компонентів лінгвістичного аналізу (наприклад, пошук відповідних слів за допомогою використання словника, здійснення синтаксичного аналізу тощо) для формування векторів ознак із подальшим застосуванням різноманітних методів машинного навчання для генерування вихідного перекладу. Відмінність моделей нейронного машинного перекладу полягає у тому, що немає попередньої екстракції різнотипних ознак; зазвичай генерування перекладу відбувається за парадигмою «Embed – Encode – Attend – Decode» (тобто, «Векторне представлення тексту – Застосування шарів енкодера – Застосування механізму уваги – Застосування шарів декодера для формування вихідних слів») [4]. Крім того, незважаючи на ефективність застосування зазначеної парадигми, компоненти статистичного машинного перекладу (наприклад, статистична

модель зіставлення слів різних мов) можуть додатково використовуватися для навчання моделі нейронного машинного перекладу [5].

Варто зазначити, що незважаючи на високу динаміку проведення досліджень у галузі машинного перекладу, залишається низка актуальних задач, що стосуються перекладу навіть поширених і схожих пар мов (наприклад, пара англійська – німецька [17]). Наприклад, одна з проблем, яку нині активно досліджують [23], полягає в здійсненні контекстно-орієнтованого перекладу [14] з урахуванням різнотипного дискурсивного зв'язку в межах всього документа (узгодження роду, формальної/неформальної стилістики, когезії тощо). Однак додатково доцільно звернути увагу на проблему *машинного перекладу усної мови*, процес здійснення якого ускладнений такими факторами:

- використання різнотипних мовленнєвих засобів, притаманних власне усній мові (наприклад, вживання слів-паразитів для формування паузи під час роздумів, коригування уже сказаних слів тощо);
- наявність помилок автоматизованого транскрибування усного мовлення.

На рис. 1 наведено приклад англійського речення, що містить водночас слово-паразит «uh» і вставну фразу «you know», а також коригування сказаних слів: «to Boston» і «to Denver». Нижче на рис. 1 подано переклади системи Google Translate на українську мову оригінального речення та версії речення, що не містить зазначених надлишкових дискурсивних одиниць.

(1) I fly you know to Boston uh to Denver.  
Я літаю, як ви знаєте, до Бостона, до Денвера.

(2) I fly to Denver.  
Я лечу в Денвер.

**Рис. 1.** Приклад здійснення перекладу системою Google Translate різних версій речення: 1 – переклад оригінальної версії речення, що містить слова-паразити та коригування слів; 2 – переклад зміненої версії речення після вилучення надлишкових слів

У разі здійснення перекладу оригінальної версії речення змінюється його семантичне значення, адже залишаються обидві фрази «до Бостона» та «до Денвера», а фраза «you know» трансформується до словосполучення «як ви знаєте». Натомість вилучення надлишкових слів дає змогу отримати коректний переклад «Я лечу в Денвер». Додатково варто зазначити складність здійснення нейронного машинного перекладу усної мови з англійської на українську через такі фактори:

- відмінність у структурному представленні речень, що ускладнює коректний переклад ситуацій коригування мови доповідачем;
- невідповідність семантичного значення слів-паразитів: якщо в англійському мовленні зазвичай використовують фрази «you know», «I mean», то для української усної мови характерні словосполучення «типу», «ну», «значить», «таки» тощо.

Отже, здійснення нейронного машинного перекладу результатів транскрибування усного мовлення з англійської мови на українську є актуальним напрямом, що потребує подальшого дослідження. Метою цієї роботи є проведення аналізу проблеми виявлення та вилучення надлишкових дискурсивних одиниць у текстах, отриманих унаслідок транскрибування усної мови, а також здійснення експериментальної перевірки ефективності застосування відповідних методів для вдосконалення нейронного машинного перекладу усного мовлення для пари англійської та української мов.

### Методи виявлення елементів дисфлюентності мовлення

*Визначення поняття дисфлюентності мовлення*

Під дисфлюентністю мовлення [10] (англ. disfluency speech) мають на увазі будь-які переривання в процесі здійснення мовлення: паузи, повторення, коригування, слова-паразити тощо. Це явище притаманне насамперед усній мові, хоча постійні регулярні переривання можуть свідчити про наявність симптомів ментального захворювання [16]. Відповідно до класифікації, наведеної в роботі [22], елементи дисфлюентності мовлення можна поділити на три категорії:

- *reparandum* – частина речення, яку було скориговано;
- *interregnum* – частина речення, що містить слова-паразити («uh», «um», «I mean», «you know» тощо);
- *repair* – частина речення, що замінює попередній *reparandum*.

На рис. 2 наведено приклад речення, що містить усі зазначені елементи.

*reparandum*
*interregnum*  
}  
*The first kind of invasion of uh I mean*  
}  
*the first type of privacy seemed invaded to me*  
}  
*repair*

**Рис. 2.** Приклад речення, що містить елементи дисфлюентності мовлення [9]

Частина речення «The first kind of invasion of» є елементом *gerarandum*, «uh I mean» – *interregnum*, «the first type of privacy» – *repair*. Виявлення слів-паразитів (*interregnum*) можливе за рахунок одночасного морфологічного (визначення частини мови для слів «uh», «um» тощо) та синтаксичного аналізу речення (відокремлення фраз «you know», «I mean» як окремих груп із подальшою їхньою класифікацією). Натомість пошук зв'язку *gerarandum*–*repair* передбачає виявлення певного типу шаблону між елементами *gerarandum* і *repair* (наприклад, елементи «The first kind of invasion» і «the first type of privacy» мають однакову структуру з відмінністю в парах слів «kind»-«type» і «invasion»-«privacy» та кінцевому слові «of»), адже коригування мовлення зазвичай здійснюється за рахунок модифікації вже сказаної фрази. Такий пошук шаблону між елементами *gerarandum* і *repair* ускладнений з таких причин:

- неможливість виявлення наведеного вище зв'язку за допомогою синтаксичного аналізатора;
- елементи *gerarandum* і *repair* можуть навіть не формувати окремі синтаксичні групи.

Отже, з метою виявлення різних елементів дисфлюентності мовлення доцільним є одночасне застосування різних методів комп'ютерної лінгвістики та машинного навчання.

#### *Сучасні методи пошуку елементів дисфлюентності мовлення*

Перші роботи [2; 12] з виявлення елементів дисфлюентності мовлення ґрунтуються на використанні різнотипних лінгвістичних властивостей слів (або ж просодичних елементів мовлення) тексту з подальшим застосуванням моделі лінійного класифікатора для формування рішення щодо типу кожної елементарної одиниці. Відповідно, пошук і вилучення пар *gerarandum*–*repair* здійснюються окремою моделлю до застосування синтаксичного парсеру. Натомість у роботах [21; 8] запропоновано виконувати пошук елементів дисфлюентності мовлення шляхом покрокового застосування синтаксичного аналізатора.

Окремо варто згадати роботу [9], в якій навчання моделі синтаксичного аналізатора, основаного на побудові конститuentного дерева (англ. *constituent parse tree*), поєднується з задачею виявлення елементів *gerarandum*, *interregnum* і *repair*. Попереднє представлення слів здійснюється за допомогою моделі BERT; такий підхід зумовлений наявністю механізму уваги в зазначеній моделі (модель BERT основана на архітек-

турі Transformer), що може допомогти виявити зв'язок між потенційними елементами *gerarandum* і *repair* упродовж процесу навчання. Зазначена модель дає змогу отримати значення F-міри, що перевищує 90 % на тестовому корпусі, для всіх типів елементів дисфлюентності мовлення [9], тож її доцільно застосовувати для фільтрування надлишкових дискурсивних одиниць, наявних в усному мовленні.

Крім того, доцільно розглянути прикладні програмні інтерфейси (API), що дають змогу здійснювати попереднє оброблення тексту з послідовним вилученням розглянутих типів елементів дисфлюентності мовлення, насамперед платформу One AI [19]. Додатково до вилучення надлишкових дискурсивних одиниць функціонал зазначеного прикладного програмного інтерфейсу дає можливість виконувати автоматизоване коригування орфографічних та пунктуаційних помилок в англійських текстах. На користь використання прикладного програмного інтерфейсу свідчить також той факт, що немає залежності від мови програмування, версій необхідних програмних бібліотек та операційної системи платформи. Враховуючи наведені переваги використання API, для здійснення експериментальної перевірки ефективності застосування методів виявлення надлишкових слів усної мови для вдосконалення нейронного машинного перекладу було використано програмний прикладний інтерфейс платформи One AI.

#### **Експериментальна частина**

##### *Розмічений корпус даних*

Метою експериментальної частини роботи є перевірка ефективності застосування методу пошуку елементів дисфлюентності мовлення для вдосконалення нейронного машинного перекладу усного мовлення для пари англійської та української мов. Першочерговим етапом є формування прикладів англійських текстів, що містять різнотипні елементи дисфлюентності мовлення: *gerarandum*, *interregnum* і *repair*. Використання зразків розміченого корпусу Switchboard [11], що застосовувався для навчання попередньо розглянутих моделей виявлення елементів дисфлюентності мовлення, ускладнене через брак відкритого безкоштовного доступу до корпусу. Тож було вирішено здійснити автоматизоване формування синтетичних прикладів дисфлюентного мовлення на основі когерентних англійських текстів природної мови. Для кожного англійського тексту  $T = \{w_1, w_1, \dots, w_N\}$ , де  $w_i, i \in \{1,$

2, ...,  $N$  – слова тексту, було створено модифікації трьох різних типів:

1. INTJ – вставка слів-паразитів «uh», «ah», «hm» у випадково обрані позиції тексту  $T$ . Кількість операцій вставлення рівна  $\lfloor N/2 \rfloor$ , де  $N$  – кількість слів тексту.
2. PRN – вставка фрази «you know» у випадково обрану позицію тексту  $T$ .
3. EDITED – вставка елементу gerarandum перед відповідним елементом гераріаі.

Із метою генерування когерентних модифікацій типу EDITED було застосовано такі послідовні кроки:

1. Формується множина слів тексту  $W \subset T$ , що належать до таких частин мови: іменник, власна назва, дієслово, займенник.
2. У випадковий спосіб здійснюється вибір слова  $w_j \in W$  зі сформованої множини.
3. Якщо слово  $w_j$  є підметом або дієсловом, виконується пошук слова  $w'_j$ , що є найближчим до слова  $w_j \in W$  у семантичному просторі:

$$w'_j = \arg \max_{w'_j \in V} (\text{sim}(w_j, w'_j)), \text{ де } V - \text{множина}$$

слів семантичного простору,  $\text{sim}(w_j, w'_j)$  – функція схожості слів (наприклад, косинусна відстань між відповідними векторами слів  $w_j$  і  $w'_j$ ). Наприклад, нехай  $w_j = \text{"cat"}$ ; тоді

$$w'_j = \arg \max_{w'_j \in V} (\text{sim}(\text{"cat"}, w'_j)) = \text{"dog"}.$$

4. Перед позицією  $j$  виконується вставка пари слів  $\langle w'_j, w_{j+1} \rangle$ . Якщо слово  $w_j \in W$  є власною назвою, додатково додається фраза “I mean” з метою формування прикладів, що містять водночас елементи коригування мовлення та слова-паразити:  $\langle w'_j, w_{j+1}, \text{"I"}, \text{"mean"} \rangle$ .

На рис. 3 наведено приклад формування синтетичних зразків дисфлюентного мовлення для різних типів модифікації речення «Frank tries to encourage Andy to learn for the tomorrow's quiz.».

Original	Frank tries to encourage Andy to learn for the tomorrow's quiz.
INTJ	Frank tries <i>um um</i> to encourage <i>um</i> Andy to learn <i>um</i> for the <i>um</i> tomorrow's quiz.
PRN	Frank tries <i>you know</i> to encourage Andy to learn for the tomorrow's quiz.
EDITED	<i>Chries tries I mean</i> Frank tries to encourage Andy to learn for the tomorrow's quiz.

Рис. 3. Формування синтетичних зразків для різних типів модифікації речення: INTJ, PRN, EDITED. Курсивом позначено сформовані елементи дисфлюентного мовлення

#### Пошук елементів дисфлюентності мовлення

Для пошуку та вилучення елементів дисфлюентності мовлення було використано при-

кладний програмний інтерфейс платформи One AI, насамперед, відповідний SDK (англ. software development kit – набір із засобів розроблення). Вхідним параметром відповідного методу SDK є текстовий документ; вихідним результатом є модифікований вхідний документ із вилученими елементами дисфлюентності мовлення. На рис. 4 наведено приклади оброблення синтетичних зразків різних типів за допомогою зазначеного SDK.

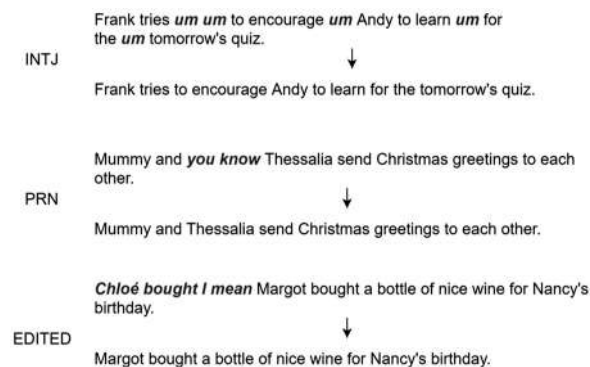


Рис. 4. Приклади оброблення синтетичних зразків дисфлюентності мовлення різних типів (INTJ, PRN, EDITED) за допомогою One AI SDK

#### Нейронний машинний переклад корпусу

Наступним етапом експериментальної частини є здійснення нейронного машинного перекладу елементів розміченого корпусу. Для кожного елемента корпусу виконується автоматизований переклад його версій, а саме:

- оригінальної версії елемента;
- синтетичних зразків елемента, що містять елементи INTJ, PRN, EDITED;
- результатів оброблення попередньо зазначених синтетичних зразків після вилучення елементів дисфлюентності мовлення.

Крім застосування наявних прикладних програмних інтерфейсів (Google Translate, DeepL Translate), можлива реалізація нейронного машинного перекладу з англійської мови на українську за допомогою попередньо навчених нейронних мереж: двомовної моделі Opus [13] та багатомовної моделі NLLB, представленої в роботі [18]. Під двомовною моделлю мається на увазі модель перекладу тексту лише з англійської мови на українську; натомість багатомовна модель передбачає можливість здійснення нейронного машинного для множини пар мов. Використання моделі Opus ускладнене нестабільністю роботи моделі: у деяких випадках вихідний переклад може формуватися російською мовою. Основною метою багатомовної моделі NLLB є

здійснення одночасного машинного перекладу для 200 різних мов; насамперед, увагу приділено перекладу текстів мов із низькою кількістю (чи навіть відсутністю) відповідних розмічених корпусів. Крім того, як зазначають автори цієї моделі, для отримання найкращої якості перекладу доцільно використовувати саме двомовну модель відповідно до вхідної пари мов [15]. Тож, урахувавши недоліки застосування зазначених моделей, здійснення автоматизованого машинного перекладу текстів з англійської мови на українську було виконано за допомогою прикладного програмного інтерфейсу системи Google Translate.

#### Програмна реалізація експериментальної частини

Програмну реалізацію експериментальної частини виконано за допомогою мови програмування Python 3.8. Для формування англійських текстів корпусу даних використано корпус SAMSum [6], а саме анотації до діалогів, написані експертами власноруч. Такий вибір зумовлено наявністю прикладів документів різної тематики (повсякденне спілкування, політика, зустрічі, чутки тощо), а також значної кількості власних назв (насамперед, імен учасників діалогу) в анотаціях, що дає змогу формувати синтетичні зразки дисфлюентності мовлення типу EDITED. Здійснення синтаксичного та морфологічного аналізу елементів корпусу виконано за допомогою бібліотеки spaCy [7]. Для виявлення найближчого слова в семантичному просторі з метою формування синтетичних зразків EDITED застосовано модель векторного представлення слів GloVe [20]. Як було зазначено раніше, вилучення надлишкових елементів дисфлюентності мовлення та здійснення машинного перекладу з англійської мови на українську було здійснено за допомогою прикладних програмних інтерфейсів Python One AI SDK та Google Translate відповідно.

#### Результати

Для кожного типу дисфлюентності мовлення (INTJ, PRN, EDITED) у випадковий спосіб було обрано 100 елементів корпусу; кожен елемент складався з оригінальної версії англійського тексту, модифікованої синтетичної версії, результату вилучення елементів дисфлюентності мовлення та відповідних українськомовних перекладів. Із метою аналізу впливу надлишкових елементів усного мовлення на якість нейронного машинного перекладу та одночасної перевірки ефективності застосування методу вилучення

елементів дисфлюентності мовлення кожен обраний елемент корпусу було класифіковано відповідно до чотирьох категорій:

- вдосконалення перекладу (*Improve*): переклад модифікованої версії із застосуванням методу вилучення елементів дисфлюентності мовлення відповідає перекладу оригінальної версії, а також є кращим, порівняно з перекладом модифікованої версії без його застосування;
- рівність перекладу (*Equal*): переклад модифікованої версії із застосуванням методу вилучення елементів дисфлюентності мовлення відповідає водночас перекладам оригінальної версії та модифікованої версії без його застосування;
- помилка перекладу (*Error*): переклад модифікованої версії із застосуванням методу вилучення елементів дисфлюентності мовлення не відповідає перекладу оригінальної версії, однак є не гіршим порівняно з перекладом модифікованої версії без його застосування;
- погіршення перекладу (*Worse*): переклад модифікованої версії із застосуванням методу вилучення елементів дисфлюентності мовлення не відповідає перекладу оригінальної версії, а також є гіршим порівняно з перекладом модифікованої версії без його застосування.

У табл. 1 наведено відсоткове співвідношення наведених вище категорій для кожного типу дисфлюентності мовлення на основі аналізу вибраної вибірки.

Таблиця 1

Відсоткове співвідношення категорій елементів вибірки корпусу (*Improve*, *Equal*, *Error*, *Worse*) для різних типів дисфлюентності мовлення

	<i>Improve</i> , %	<i>Equal</i> , %	<i>Error</i> , %	<i>Worse</i> , %
INTJ	65.66	19.19	12.12	3.03
PRN	71.28	20.79	6.93	1.00
EDITED	62.37	11.88	19.80	5.95

Наведені вище результати свідчать про доцільність застосування використаного методу вилучення елементів дисфлюентності мовлення для вдосконалення якості нейронного машинного перекладу з англійської мови на українську: для всіх типів дисфлюентності мовлення відсоткове значення елементів категорії вдосконалення перекладу *Improve* перевищує 60 %, а відповідне значення для категорії погіршення перекладу *Worse* становить менше ніж 6 %. Значення *Equal* = 19.19 % та *Equal* = 20.79 % для типів INTJ та PRN відповідно вказують на спроможність моделі нейронного машинного перекладу ігнорувати та вилучати певні поодинокі сло-

ва-паразити («uh», «ah», «hm», «you know») протягом формування вихідного тексту. Однак наявність багатьох таких послідовних слів призводить до спотворення перекладу. Наприклад, розглянемо такі пари «оригінальна версія – переклад» і «модифікована версія – переклад»:

- «Serena's skin condition is fine now and she doesn't have to take medication.» – «Стан шкіри Серени зараз нормальний, і їй не потрібно приймати ліки.»
- «Serena's ah skin condition is fine now and ah she doesn't ah have ah ah ah ah ah ah to ah take medication ah.» – «Зараз стан шкіри ах Серени в порядку, і ах вона не ах ах ах ах ах ах ах ах приймати ліки ах.»

Крім ускладнення сприйняття перекладеного речення через наявність надлишкових слів, варто зазначити про помилку в узгодженості форми дієслова «приймати» з відповідним займенником «вона», що додатково вказує на доцільність вилучення слів-паразитів такого типу.

Сумарне відношення кількості правильних перекладів, отриманих за допомогою застосування методу вилучення елементів дисфлюентності мовлення, до їх загальної кількості становить 84.85 %, 92.07 % і 74.25 % для типів INTJ, PRN і EDITED відповідно. Найнижче значення отримано для типу EDITED, що підтверджує складність виявлення зв'язку *gerarandum-gerair* у межах речення. Крім того, значення відсотка категорії *Worse* = 5.95 % для цього типу є вищим порівняно з типами INTJ і PRN. Така відмінність полягає у помилковому вилученні слів, класифікованих як *gerarandum*, що призводить до модифікації інформації, поданої в реченні. Наприклад, розглянемо переклади для оригінальної версії речення, модифікованої версії та версії, отриманої внаслідок застосування методу вилучення елементів дисфлюентності мовлення:

- «Nathalie, Pauline, Jacob and Anthony are thinking about spending holidays in Greece together in August.» – «Наталі, Поліна, Джейкоб і Ентоні планують разом провести відпустку в Греції в серпні.»
- «*Lona, I mean* Nathalie, Pauline, Jacob and Anthony are thinking about spending holidays in Greece together in August.» – «Лона, я маю на увазі Наталі, Полін, Джейкоб і Ентоні думають про те, щоб провести разом відпустку в Греції в серпні.»
- «*Lona*, Nathalie, Pauline, Jacob and Anthony are thinking about spending holidays in Greece together in August.» – «Лона, Наталі, Поліна, Джейкоб і Ентоні думають про те, щоб разом відпочити в Греції в серпні.»

Елементом *gerarandum* у модифікованому реченні є фраза «Lona, I mean». Однак вилучення лише слів «I mean», здійснене за допомогою розглянутого методу, призводить до некоректного трактування переліку осіб на початку речення, що в подальшому впливає на відповідний переклад. Отже, врахування впливу такого вилучення слів на зміну семантичного значення речення є актуальною проблемою і потребує подальших досліджень.

## Висновки

У роботі проаналізовано проблеми виявлення та вилучення надлишкових дискурсивних одиниць у текстах, отриманих унаслідок транскрибування усної мови, а також здійснено експериментальну перевірку ефективності застосування відповідних методів пошуку елементів дисфлюентності мовлення для вдосконалення нейронного машинного перекладу усного мовлення для пари англійської та української мов. Враховуючи проведений аналіз впливу елементів дисфлюентності мовлення на кінцевий українськомовний переклад та отримані результати класифікації сформованих перекладів, можна зробити такі висновки:

- Зважаючи на вплив наявності елементів дисфлюентності мовлення на семантичне значення кінцевого українськомовного перекладу, актуальним завданням є виявлення та вилучення відповідних дискурсивних одиниць, насамперед зв'язку *gerarandum-gerair*, що інтерпретує коригування думок доповідача протягом його мовлення.
- Аналіз перекладу речень, що містять слова-паразити усного мовлення, вказує на можливість водночас вилучення відповідних слів, а також здійснення їхнього перекладу, що не збігається з українськомовними відповідниками, під час застосування моделі нейронного машинного перекладу. Тож із метою збереження стилістики оригінального речення актуальним напрямом є дослідження здійснення нейронного машинного перекладу з урахуванням слів-паразитів в українськомовному кінцевому тексті відповідно до особливостей їх використання в українській мові.
- Із метою вдосконалення якості нейронного машинного перекладу англійської усної мови на українську доцільним є застосування методу вилучення елементів дисфлюентності мовлення, що підтверджують отримані значення відношення кількості коректних перекладів на множині різнотипних модифікацій англомовних речень.

## Список літератури

- Bahdanau D. Neural Machine Translation by Jointly Learning to Align and Translate / D. Bahdanau, K. Cho, Y. Bengio. – 2014.
- Charniak E. Edit Detection and Parsing for Transcribed Speech / E. Charniak, M. Johnson // Second Meeting of the North American Chapter of the Association for Computational Linguistics. – 2001.
- Cho K. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches / K. Cho, B. van Merriënboer, D. Bahdanau, Y. Bengio // Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. – Stroudsburg, PA, USA : Association for Computational Linguistics, 2014. – Pp. 103–111.
- Dabre R. A Survey of Multilingual Neural Machine Translation / R. Dabre, C. Chu, A. Kunchukuttan // ACM Comput. Surv. – 2020. – Vol. 53, No. 5. – Pp. 1–38.
- Garg S. Jointly Learning to Align and Translate with Transformer Models / S. Garg, S. Peitz, U. Nallasamy, M. Paulik // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). – Hong Kong, China : Association for Computational Linguistics, 2019. – Pp. 4453–4462.
- Gliwa B. SAMSum Corpus: A Human-annotated Dialogue Dataset for Abstractive Summarization / B. Gliwa, I. Mochol, M. Biesek, A. Wawer // Proceedings of the 2nd Workshop on New Frontiers in Summarization. – Hong Kong, China : Association for Computational Linguistics, 2019. – Pp. 70–79.
- Honnibal Matthew. spaCy – Industrial-strength Natural Language Processing in Python [Electronic resource] / Honnibal Matthew, Montani Ines. – 2015. – Mode of access: <https://spacy.io>.
- Honnibal M. Joint Incremental Disfluency Detection and Dependency Parsing / M. Honnibal, M. Johnson // Transactions of the Association for Computational Linguistics. – 2014. – Vol. 2. – Pp. 131–142.
- Jamshid Lou P. Improving Disfluency Detection by Self-Training a Self-Attentive Model / P. Jamshid Lou, M. Johnson // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). – Online : Association for Computational Linguistics, 2020. – Pp. 2756–2765.
- Jamshid Lou P. Neural Constituency Parsing of Speech Transcripts / P. Jamshid Lou, Y. Wang, M. Johnson // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. – Minneapolis, Minnesota : Association for Computational Linguistics, 2019. – Pp. 3754–3763.
- John J. Godfrey Switchboard-1 Release 2 [Electronic resource] / John J. Godfrey, Edward Holliman. – 1997. – Mode of access: <https://catalog.ldc.upenn.edu/LDC97S62>.
- Kahn J. G. Effective Use of Prosody in Parsing Conversational Speech / J. G. Kahn, M. Lease, E. Charniak et al. // Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. – Vancouver, British Columbia, Canada : Association for Computational Linguistics, 2005. – Pp. 233–240
- Language Technology Research Group at the University of Helsinki Helsinki-NLP/opus-mt-en-uk · Hugging Face [Electronic resource] / Language Technology Research Group at the University of Helsinki. – 2020. – Mode of access: <https://huggingface.co/Helsinki-NLP/opus-mt-en-uk>.
- Maruf S. A Survey on Document-Level Neural Machine Translation: Methods and Evaluation / S. Maruf, F. Saleh, G. Haf-fari // ACM Comput. Surv. – 2021. – Vol. 54, No. 2.
- Meta AI. 200 languages within a single AI model: A breakthrough in high-quality machine translation [Electronic resource]. – 2022. – Mode of access: <https://ai.facebook.com/blog/nllb-200-high-quality-machine-translation/>.
- Miltenberger Raymond G. Speech Disfluencies / Raymond G. Miltenberger, D. W. Woods // Handbook of Child Behavior Therapy (pp. 127–142 / S. T. Watson and F. M. Gresham. – Boston, MA : Springer US, 1998. – Pp. 127–142.
- Müller M. A Large-Scale Test Set for the Evaluation of Context-Aware Pronoun Translation in Neural Machine Translation / M. Müller, A. Rios, E. Voita, R. Sennrich // Proceedings of the Third Conference on Machine Translation: Research Papers. – Brussels, Belgium : Association for Computational Linguistics, 2018. – Pp. 61–72.
- NLLB Team No Language Left Behind: Scaling Human-Centered Machine Translation / NLLB Team, M. R. Costa-jussà, J. Cross et al. – 2022.
- One AI Inc. One AI Language Studio [Electronic resource]. – 2022. – Mode of access: <https://studio.oneai.com>.
- Pennington J. GloVe: Global Vectors for Word Representation / J. Pennington, R. Socher, C. Manning // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). – Doha, Qatar : Association for Computational Linguistics, 2014. – Pp. 1532–1543.
- Rasooli M. S. Joint Parsing and Disfluency Detection in Linear Time / M. S. Rasooli, J. Tetreault // Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. – Seattle, Washington, USA : Association for Computational Linguistics, 2013. – Pp. 124–129.
- Shriberg E. Preliminaries to a Theory of Speech Disfluencies / E. Shriberg. – 1994.
- Sun Z. Rethinking Document-level Neural Machine Translation / Z. Sun, M. Wang, H. Zhou et al. // Findings of the Association for Computational Linguistics: ACL 2022. – Dublin, Ireland : Association for Computational Linguistics, 2022. – Pp. 3537–3548.
- Wu Y. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation / Y. Wu, M. Schuster, Z. Chen et al. – 2016.

## References

- Bahdanau, D., Cho, K., & Bengio, Y. (2014). *Neural Machine Translation by Jointly Learning to Align and Translate*.
- Charniak, E., & Johnson, M. (2001). Edit Detection and Parsing for Transcribed Speech. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*. <https://aclanthology.org/N01-1016>.
- Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111. Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-4012>
- Dabre, R., Chu, C., & Kunchukuttan, A. (2020). A Survey of Multilingual Neural Machine Translation. *ACM Computing Surveys*, 53 (5), 1–38. <https://doi.org/10.1145/3406095>
- Garg, S., Peitz, S., Nallasamy, U., & Paulik, M. (2019). Jointly Learning to Align and Translate with Transformer Models. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 4453–4462. <https://doi.org/10.18653/v1/D19-1453>
- Gliwa, B., Mochol, I., Biesek, M., & Wawer, A. (2019). SAMSum Corpus: A Human-annotated Dialogue Dataset for Abstractive Summarization. In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, 70–79. <https://doi.org/10.18653/v1/D19-5409>
- Honnibal Matthew, & Montani Ines. (2015). *spaCy – Industrial-strength Natural Language Processing in Python*. <https://spacy.io/>.
- Honnibal, M., & Johnson, M. (2014). Joint Incremental Disfluency Detection and Dependency Parsing. *Transactions of the*

- Association for Computational Linguistics*, 2, 131–142. [https://doi.org/10.1162/tacl\\_a\\_00171](https://doi.org/10.1162/tacl_a_00171)
- Jamshid Lou, P., & Johnson, M. (2020). Improving Disfluency Detection by Self-Training a Self-Attentive Model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 3754–3763. <https://doi.org/10.18653/v1/2020.acl-main.346>
- Jamshid Lou, P., Wang, Y., & Johnson, M. (2019). Neural Constituency Parsing of Speech Transcripts. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2756–2765. <https://doi.org/10.18653/v1/N19-1282>
- John J. Godfrey, & Edward Holliman. (1997). *Switchboard-1 Release 2*. <https://catalog.ldc.upenn.edu/LDC97S62>.
- Kahn, J. G., Lease, M., Charniak, E., Johnson, M., & Ostendorf, M. (2005). Effective Use of Prosody in Parsing Conversational Speech. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 233–240. <https://aclanthology.org/H05-1030>.
- Language Technology Research Group at the University of Helsinki. (2020). *Helsinki-NLP/opus-mt-en-uk · Hugging Face*. <https://huggingface.co/Helsinki-NLP/opus-mt-en-uk>.
- Maruf, S., Saleh, F., & Haffari, G. (2021). A Survey on Document-Level Neural Machine Translation: Methods and Evaluation. *ACM Comput. Surv.*, 54 (2). <https://doi.org/10.1145/3441691>
- Meta AI. (2022). 200 languages within a single AI model: A breakthrough in high-quality machine translation. <https://ai.facebook.com/blog/nllb-200-high-quality-machine-translation/>.
- Miltenberger, Raymond G., & Woods, D. W. (1998). Speech Disfluencies. In F. M. S. T. Watson & F. M. Gresham (Eds.), *Handbook of Child Behavior Therapy* (pp. 127–142). Springer US. [https://doi.org/10.1007/978-1-4615-5323-6\\_7](https://doi.org/10.1007/978-1-4615-5323-6_7)
- Müller, M., Rios, A., Voita, E., & Sennrich, R. (2018). A Large-Scale Test Set for the Evaluation of Context-Aware Pronoun Translation in Neural Machine Translation. *Proceedings of the Third Conference on Machine Translation: Research Papers*, 61–72. <https://doi.org/10.18653/v1/W18-6307>
- NLLB Team, Costa-jussà, M. R., Cross, J., Çelebi, O., Elbayad, M., Heafield, K., Heffernan, K., Kalbassi, E., Lam, J., Licht, D., Maillard, J., Sun, A., Wang, S., Wenzek, G., Youngblood, A., Akula, B., Barrault, L., Gonzalez, G. M., Hansanti, P., ... Wang, J. (2022). *No Language Left Behind: Scaling Human-Centered Machine Translation*.
- One AI Inc. (2022). One AI Language Studio. <https://studio.oneai.com>.
- Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. <https://doi.org/10.3115/v1/D14-1162>
- Rasooli, M. S., & Tetreault, J. (2013). Joint Parsing and Disfluency Detection in Linear Time. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 124–129. Association for Computational Linguistics. <https://aclanthology.org/D13-1013>
- Shriberg, E. (1994). *Preliminaries to a Theory of Speech Disfluencies*.
- Sun, Z., Wang, M., Zhou, H., Zhao, C., Huang, S., Chen, J., & Li, L. (2022). Rethinking Document-level Neural Machine Translation. *Findings of the Association for Computational Linguistics: ACL 2022*, 3537–3548. <https://doi.org/10.18653/v1/2022.findings-acl.279>
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., ... Dean, J. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*.

A. Kramov, S. Pogorilyy

## USAGE OF THE SPEECH DISFLUENCY DETECTION METHOD FOR THE MACHINE TRANSLATION OF THE TRANSCRIPTIONS OF SPOKEN LANGUAGE

*Neural machine translation falls into the category of natural language processing tasks. Despite the availability of a big number of research papers that are devoted to the improvement of the quality of the machine translation of documents, the problem of the translation of the spoken language that contains the elements of the disfluency speech is still an actual task, especially for low-resource languages like the Ukrainian language. In this paper, the problem of the neural machine translation of the transcription results of the spoken language that incorporate different elements of the disfluency speech has been considered in the case of the translation from the English language to the Ukrainian language. Different methods and software libraries for the detection of the elements of disfluency speech in English texts have been analyzed. Due to the lack of open-access corpora of the speech disfluency samples, a new synthetic labeled corpus has been created. The created corpus contains both the original version of a document and its modified version according to the different types of speech disfluency: filler words (uh, ah, etc.) and phrases (you know, I mean), reparandum-repair pairs (cases when a speaker corrects himself during the speech). The experimental verification of the effectiveness of the usage of the method of disfluency speech detection for the improvement of the machine translation of the spoken language has been performed for the pair of English and Ukrainian languages. It has been shown that the current state-of-the-art neural translation models cannot produce the appropriate translation of the elements of speech disfluency, especially, in the reparandum-repair cases. The results obtained may indicate that the mentioned method of disfluency speech detection can be used for the previous processing of the transcriptions of spoken dialogues for the creation of coherent translations by the usage of the different models of neural machine translation.*

**Keywords:** natural language processing, neural machine translation, speech disfluency, processing of spoken language.

