

*О. Д. Лагута*

*Житомирський державний університет*

*імені І. Франка*

*Науковий керівник:*

*к.ф.н., доц. Гирич О. В.*

## **ПРАКТИЧНИЙ ПОТЕНЦІАЛ СИНТАКСИЧНИХ АНАЛІЗАТОРІВ РОЗПІЗНАВАТИ ТА КОРЕКТУВАТИ ГРАМАТИЧНО НЕПРАВИЛЬНІ РЕЧЕННЯ АНГЛІЙСЬКОЮ МОВОЮ**

**Вступ.** Автоматична обробка природної широко використовується у лінгвістиці, а отже, автоматичний синтаксичний аналіз є найважливішою частиною цього процесу, тобто являє собою основу, необхідною для поглибленого аналізу.

Автоматичний синтаксичний аналіз речення використовується, щоб визначити “синтаксичне дерево” (parsing tree), з метою зробити граматичний аналіз речення, тобто розкрити його граматичну структуру. Синтаксичне дерево - це упорядковане кореневе дерево, яке являє собою синтаксичну структуру рядка (речення) згідно певної граматики [1, р. 34].

Як вже було зазначено вище, автоматичний синтаксичний аналіз (парсинг) є основою обробки природної мови, тому, як і будь-яке наукове явище, парсинг має свою сферу використання та завдання, які повинен виконувати. Одним з найважливіших завдань парсингу є перевірка граматики. Характерним показником даного способу перевірки граматики є те, що програма, проаналізувавши текст, може виділити наявні граматичні помилки в реченні, але не завжди зможе повідомити користувачеві, якого роду ця помилка. Для цього програма потребує додаткових правил, які можуть аналізувати граматично неправильні речення.

**Мета** даної статті полягає у тому, щоб на прикладі двох конкретних програм, завданням яких є перевірка граматики (“GINGER” Grammar Checker [5] and “Language Tool[4]) перевірити їх можливості розпізнавання та виправлення граматичних помилок в англійському реченні. Обидві програми є безкоштовними, вони можуть виділити помилки, запропонувати іншу альтернативу написання речення, але також існують додаткові функції глибокого розбору речення, які будуть доступними лише після їх купівлі. Автори даних програм стверджують, що вони здатні перевірити помилки майже будь-якого типу, тому, перш за все, варто дати характеристику найпоширенішим граматичним помилкам, які зустрічаються в англійській мові. Граматичні помилки можна розділити на дві групи: структурні та не структурні. Структурні помилки – це помилки, які можна виправити лише шляхом вставки, видалення або переміщення одного або декількох слів, а не структурні помилки – це ті, які можна виправити шляхом заміни існуючого слова іншим[2].

Розглянемо декілька прикладів типових помилок та результати використання вище зазначених програм:

**Неправильне узгодження підмета з дієсловом** (Subject-Verb Agreement).

Підмет і дієслово повинні узгоджуватися між собою, залежно від того, в якому числі вони стоять – в однині чи в множині [3]

Дієслово-присудок повинно узгоджуватися в числі з підметом речення: якщо підмет стоїть в однині, то й дієслово-присудок також має бути в однині, а множина підмета зумовлює й множину дієслова-присудка [2]. Наприклад:

**Неправильно:** *The benefits of using this program is underestimated.*

**Правильно:** *The benefits of using this program are underestimated.*

Іменник “*The benefits*” стоїть у множині, тому, відповідно і дієслово повинно бути у множині. Очевидно, що носії англійської мови або люди, які добре знають граматику англійської мови, зможуть легко розпізнати помилку в цьому реченні. Пропонуємо розглянути результати перевірки граматики програмою “GINGER”. Ввівши наше речення, програма, через декілька секунд видала наступний результат:

*The benefits of using this program is underestimated* (“is underestimated” підкреслено, а нижче був наведений виправлений варіант речення: *The benefits of using this program **are underestimated***. Отже, в даному випадку, програма зреагувала на наявність неузгодження підмета з дієсловом, вказала на помилку та виправила її.

Ввівши вище запропоноване речення у рядок запиту (query string), програма “LanguageTool” не змогла розпізнати помилку.

**Часова форма дієслова (Verb Tense).**

Дієслово, використане у неправильній часовій формі може змінити сенс речення, що призведе до труднощів розуміння та інтерпретування речення читачем або співрозмовником. Часова форма дієслова має розповідати про те, коли відбувається дія – в минулому, майбутньому чи теперішньому. Розглянемо приклад граматично неправильного речення:

**Неправильно:** *Our last meeting take place yesterday.*

**Правильно:** *Our last meeting took place yesterday..*

Результати перевірки за допомогою програми “GINGER”:

Як і в попередньому випадку, дана програма вказала на помилку:

*Our last meeting take place yesterday*, та виправила її: *Our last meeting **took** place yesterday.*

Очевидно, що програма зреагувала на наявність прислівника “yesterday”, що є словом-маркером у використанні простого минулого часу (Past Indefinite Tense) або прикметник “last”.

Результати перевірки за допомогою програми “Language Tool”:

Запропоноване речення залишилось без виправлень, тобто програма не змогла ідентифікувати наявні помилки, що означає, що, в процесі перевірки часової форми дієслова, алгоритм програми не орієнтується на слова-маркери використання окремих часових форм дієслів англійської мови.

**Число іменників (Noun Number).**

До помилок даного типу відносяться помилки, пов'язані неправильним вживанням числа іменника (залежно від контексту, в однині чи в множині), що робить їх доволі поширеним явищем. Такі помилки є характерними для людей, які вивчають англійську мову як іноземну, тобто не є її носіями.

Керуючись загальними правилами утворення множини іменників в англійській мові, люди роблять помилки, якщо справа стосується винятків, тобто тих іменників, до формування множини яких не застосовуються загальні правила.

Наведемо простий приклад:

**Неправильно:** *How many person have you invited?*

**Правильно:** *How many people have you invited?*

Результати перевірки за допомогою програми “GINGER”:

Програма успішно розпізнала помилку та виправила її:

*How many person have you invited? - How many **people** have you invited?*

Дивлячись на це, можемо зробити припущення, що дана програма зреагувала на наявність перед іменником “person” прислівника “many”, що, в свою чергу може означати, що програма керувалась правилами застосування питального речення “How many...”, де “many” вживається перед іменниками, які стоять у множині і які можна порахувати (countable nouns). Підтвердженням цього припущення може бути наявність запропонованих програмою інших альтернатив використання даної структури: *How many people have you impacted? How many people have you helped?*

Результати перевірки за допомогою програми “Language Tool”:

Дана програма розпізнала помилку даного характеру, виділивши слова:

*How **many person** have you invited?*

Крім того, що програмою було розпізнано помилку, програма також вказала на можливий характер помилки: “Possible agreement error. The noun seems to be countable – consider using “many people”. Очевидно, що ця програма керувалась тими ж самими правилами, що і попередня, але на відміну від “GINGER”, ця програма змогла виявити те тільки саму помилку, але й вказати на її характер та запропонувала правильний варіант.

### **Порушення правил використання омонімів.**

Деякі пари чи групи слів є незрозумілими, оскільки вони схожі, але мають різні значення. Наприклад:

**It's/Its.**

*Its going to be a long day. Does the car need it's oil changed?*

У даному прикладі “Its” це присвійний займенник, а “it's” є скороченою формою від “it is”.

Результати перевірки за допомогою програми “GINGER”:

Запропоноване речення: *Its going to be a long day. Does the car need it's oil changed?*

Виправлене програмою речення: *It's going to be a long day. Does the car need its oil changed?*

Отже, так як і в попередніх випадках, програма успішно розпізнала граматичні помилки, виправила їх, але не вказала на їх характер. Варто

додати, що, не дивлячись на те, що програма не вказала на характер помилок, нею було запропоновано інші альтернативи, але тільки стосовно “It’s”:

*It's going to be a long day again. It's going to be a long day folks.*

Можна зробити висновок, що програма орієнтувалася на правила застосування структури “to be going to”, тому, розпізнавши слово “going”, але побачивши перед ним не “It’s” (to be in Present Tense), а присвійний займенник “its”, зреагувала на помилку та виправила її. Щодо другого речення, то тут також все зрозуміло. Очевидно, згідно алгоритму, всі присвійні займенники обов’язково повинні бути «прив’язаними» до іменника.

Результати перевірки за допомогою програми “Language Tool”:

Виділені програмою помилки: *Its going to be a long day. Does the car need it's oil changed?* Навівши комп’ютерною мишкою на ці дві помилки, з’явилися «вікна» з питаннями:

*“Did you mean “it’s” (= “it is”) instead of “its” (possessive pronoun)?*

*“Did you mean “its” (possessive pronoun) instead of “it’s” (= “it is”)?”*

Як ми бачимо, програма розпізнала помилки та частини мови, які є у реченні. Окрім цього, програма також наводить декілька прикладів використання омонімів “its” та “it’s”, проілюструємо один з них:

**Incorrect:** *For all it's perfection, I could not really like it.*

**Correct:** *For all its perfection, I could not really like it.*

**Висновки.** Отже, як ми бачимо, безкоштовний режим користування програмами дає можливість зробити синтаксичний аналіз речень та розпізнати граматичні помилки. Розглянувши декілька прикладів граматично неправильних речень та перевіривши здатність програм розпізнавати та виправляти помилки, можемо зробити висновок, що обидві програми мають свої переваги та недоліки, та, більшою мірою, керуються схожими правилами перевірки та коректування помилок. Очевидно, що в «арсеналі» програми існує багато прикладів використання різних структур, та у випадку, якщо хоч один з елементів випадає з цією структури, програма реагує на це неузгодження. Програми не є ідеальними, але свою функцію, хоч і не повномірно, виконують. Також очевидним є те, що отримання таких результатів було б неможливими без використання парсеру (parser), тобто синтаксичного аналізу, додатком якого має бути синтаксичне дерево. При отриманні результатів перевірки граматики, користувачеві не демонструється синтаксичне дерево (parsing tree), а одразу, в випадку успішної перевірки, наявність помилок та виправлене речення.

Список використаної літератури:

1. Chapman, N. P. LR parsing: Theory and Practice / N. P. Chapman. – NY.: Cambridge University Press, 1987. – 211 p.
2. Frost J. The 12 Most Common Grammar Errors [Електронний ресурс] / J. Frost. Режим доступа: <https://www.grammarcheck.net/the-12-twelve-most-common-grammar-errors/>
3. Linville C. Common Error Types for English Language Learners [Електронний ресурс] / C. Linville. Режим доступа: [https://www.sbccc.edu/clrc/writing\\_center/Common%20Error%20Types%20for%20English%20Language%20Learners.pdf](https://www.sbccc.edu/clrc/writing_center/Common%20Error%20Types%20for%20English%20Language%20Learners.pdf)

4. Naber D. Language Tool [Электронный ресурс] / D. Naber. Режим доступа: <https://languagetool.org/>
5. Sacht M. Ginger Grammar Checker [Электронный ресурс] / М. Shacht. Режим доступа: <https://www.gingersoftware.com/grammarcheck#.W-x3A-gzbIU>