

4. Болюбаш Я. Я. Організація навчального процесу у вищих закладах освіти. К.: Наукова думка, 1997. С. 62.

5. Бондар В. І. Дидактика: ефективні технології навчання студентів. К.: "Вересень", 1996. 129 с.

6. Вакуленко В. М. Основи вищої школи України. Навчальний посібник. Луганськ: вид-во СНУ, 2001. 247 с.

*Юценко Світлана Сергіївна,*

науковий співробітник Українського мовно-інформаційного фонду НАН України

### **АВТОРОЗНАВЧА ЕКСПЕРТИЗА ЗА УМОВ ПОВНОЇ НЕВИЗНАЧЕНОСТІ**

Авторознавча експертиза – це комплексне лінгвістичне дослідження письмового тексту на предмет встановлення фактичних даних про особу автора. Коло завдань, що вирішує авторознавча експертиза, є широким; зупинимося на ідентифікаційних, розв’язання яких розуміємо як однозначне науково обгрунтоване атрибутування аналізованого тексту конкретній особі.

Для атрибуції підписаного тексту необхідний порівняльний матеріал – достовірні зразки тексту особи, авторство якої підлягає підтвердженню чи спростуванню, анонімного – зразки тексту особи, яка, за припущенням ініціатора експертизи або експертів, може бути його автором.

Зразки обсягу, недостатнього для виявлення індивідуальних рис авторського стилю, або сумнівні зразки, або відсутність зразків без можливості отримання називаємо «умовами повної невизначеності».

На сьогодні методики такого роду досліджень не існує, ми виносимо на обговорення власну, апробовану на текстах реальної складності статистично репрезентативного обсягу (кілька сотень сторінок), що передбачає застосування програмних мовно-орієнтованих інструментів, розроблених в Українському мовно-інформаційному фонді НАН України (далі – Фонд) – Когнітивної ІТ-платформи «ПОЛІЕДР» (у співпраці з Інститутом кібернетики ім. В. М. Глушкова НАН України і Національним центром «Мала академія наук України»)¹ та Системи аналізу текстів (CAT) [1].

---

¹ Свідцтво про авторське право на твір № 96078 від 17 лютого 2020 року “Комп’ютерна програма “Когнітивна ІТ платформа ПОЛІЕДР” (“КІТ ПОЛІЕДР”) (“POLYHEDRON”). Автори: Стрижак О.С., Глоба Л. С., Величко В. Ю., Попова М. А., Широков В. А., Надутенко М. В. та ін. Офіційний бюлетень № 57 (від 31 березня 2020 року), с. 402 – 403.

Згідно з нашою методикою, на початковому етапі авторознавчої експертизи з метою звуження кола ймовірних авторів вивчаються обставини (часові, просторові, об'єктні, суб'єктні тощо) створення та функціонування винесеного на експертний розгляд тексту (далі – аналізований текст), відомі та встановлені безпосередньо з нього.

Наступним етапом є окреслення предмету дослідження, що полягає у вилученні цитат та запозичених текстів інших авторів для зменшення впливу на результати досліджень. Після чого кожне речення редукованого в такий спосіб тексту надсилається на пошукові системи для встановлення дублікатів серед текстів вебу. В окремих випадках цього буває достатньо для атрибуції аналізованого тексту, оскільки, за нашими спостереженнями, переважна більшість авторів вдаються до самоцитування.

За результатами вебпошуку формується аналітичний масив, комп'ютерне порівняння текстів якого з аналізованим дозволяє атрибутувати його певну частину, оскільки, знов-таки за нашими спостереженнями, якщо текст створений особою, недоброчесною як автор, обов'язково відшукається компіляція. За допомогою антиплагіатної програми (як зазначалося, Фонд послуговується програмним продуктом власної розробки – Системою аналізу текстів) визначаються та вилучаються як атрибутовані, фрагменти, що є дублікатами оприлюднених текстів. Об'єктом подальшого дослідження є редукований в описаний спосіб текст – без цитат та запозичень.

Власне кажучи, атрибуція текстів можлива завдяки зв'язку між мовними структурами та структурами мислення – цей фундаментальний факт доволі твердо встановлено наукою. Кожна людина, крім того, що володіє унікальним (суб'єктивним) лексиконом, сформованим набутими протягом життя знаннями та досвідом, також володіє правилами оперування цим лексиконом. Хоча суб'єктивний лексикон фактично необмежений, в ньому обов'язково є зона індивідуальних лексичних домінант. Опис будь-якого предмету думки автор здійснює, використовуючи, насамперед, лексику цієї зони.

Отже, при проведенні авторознавчої експертизи пошуково-дослідницька діяльність експертів спрямовується на визначення та опис індивідуальних особливостей стилю фактичного автора аналізованого тексту, до яких, зокрема належать: помилки писемної комунікації (правописні, пунктуаційні, синтаксичні, семантичні, логічні тощо); слова-маркери та слова-паразити; рідковживана, застаріла, просторічна, жаргонна, арготична, діалектична, іншомовна, поетична, функціональна (службова) та ін. лексика; наявність та

особливості подання цитат, жартів, звернень, вставних конструкцій, дат, проміжків часу, власних назв та ін.

Також предметом уваги експертів є особливості пунктуації та синтаксичної будови речень, стилістичні прийоми, комунікативна мета, читацька адреса, фонові знання та ін.

Наступним етапом авторознавчої експертизи є статистичні дослідження, що дозволяють виявляти неконтрольовані параметри писемного мовлення (частотність та розподіл частин мови, ключових та функціональних (службових) слів; речень та слів за довжиною та інші). На цьому етапі наша методика передбачає також пошук та аналіз розподілу колокацій (N-грам, тобто ланцюгів (послідовностей) з n слів, де  $n = 1, 2, 3 \dots$ ), які, на відміну від художньої, рідковживаної та іншої специфічної лексики, наявні в кожному тексті будь-якого функціонального стилю. Приклади N-грам довжини 2 (*дозволимо собі; відверто кажучи; отже, ми; особливо гостро*); довжини 3 (*довести до кінця; розглянемо більш детально; як слушно зауважує*); довжини 4 (*як кажуть у нас; є всі підстави вважати*), довжини 5 (*і це ще не все*).

N-грами фактично є мовленнєвими штампами, стереотипно відтворюваними регулярними синтаксичними конструкціями – у художніх текстах для створення певного художнього образу, надання індивідуальності мовленню діючим особам (здаємо асоційовані з літературними персонажами або публічними особами слова та фрази). В офіційних – навпаки, штампи слугують для маскування недоречної творчої індивідуальності. У публіцистичних текстах штампи забезпечують економію мовленнєвих зусиль та «готову художність» (*локомотив розвитку; острівець безпеки; усякими правдами і неправдами; історія свідчить*), слугують для створення та закріплення «потрібних» автору стереотипних образів.

Із накопиченням фактів експерти Українського мовно-інформаційного фонду дійшли висновку щодо не випадкового характеру N-грам множинного (більше одного) входження, що дозволяє також розцінювати їх як складники індивідуального способу (стилю) мовлення.

Отримані результати характеризують індивідуальний стиль (індивідуальні писемно-мовленнєві навички) фактичного автора аналізованого тексту, його особистість, світогляд, свідомість, компетенцію.

Далі згідно із запропонованою методикою здійснюється багатопараметричний пошук фрагментів із подібним проявом мовленнєвої особистості автора в неатрибутованій частині аналізованого тексту та в веб. Пошуковими параметрами виступають

визначені експертами складники індивідуального стилю у сукупності. Як зазначалося у вступній частині, експерти Фонду виконують цей пошук за допомогою Когнітивної ІТ-платформи «ПОЛІЄДР». Звісно, шукати можливо в ручному режимі, але, на нашу думку, трудомісткість такої роботи для середніх та великих за обсягом текстів перевищуватиме межі розумного.

Статистично тексти природної мови, як лінгвістичні феномени, представлені комбінаціями великої кількості доволі рідкісних подій, тому в атрибуційних дослідженнях збіг сукупності лінгвістичних фактів вважається малоімовірним. Тобто, з урахуванням встановлених обставин створення аналізованого тексту, автор із найбільшим збігом параметрів індивідуального стилю з параметрами, встановленими в аналізованому тексті, із високою ймовірністю є його фактичним автором

Із зазначеного випливає, що обов'язковою умовою застосовності розглянутої методики є наявність результатів літературної роботи фактичного автора аналізованого тексту в вебі (як варіант – у локальній базі, сформованій на початок експертного дослідження з тематичних текстів або текстів ймовірних авторів). Наведене демонструє обмеженість можливостей авторознавчої експертизи загалом – професіонально (або навіть винахідливо) виконану компіляцію атрибутувати практично неможливо.

#### **Список використаних джерел:**

1. Широков В. А., Надутенко М. В., Стрижак О. Є., Ющенко С. С. Технологічні засади логіко-лінгвістичних досліджень законодавства. Науково-технічний журнал «Біоніка інтелекту». Том 2. № 95 (2020). С. 3 – 14. (режим доступу: <https://journals.indexcopernicus.com/api/file/viewByFileId/1216174.pdf>).