

## ВІДГУК ОФІЦІЙНОГО ОПОНЕНТА

доктора технічних наук, професора

**Лупенка Сергія Анатолійовича**

на дисертаційну роботу **Хомицької Ірини Юріївни**

«Методи та засоби диференціації фоностатистичних структур  
функціональних стилів англійської мови»

поданої на здобуття наукового ступеня кандидата технічних наук  
за спеціальністю 10.02.21 – структурна, прикладна та математична  
лінгвістика

*Актуальність теми.* Авторка дисертаційного дослідження дає обґрунтування важливості розв'язання завдання підвищення достовірності встановлення авторства тексту. Зростаюче зацікавлення цим напрямом дослідження зумовлене широкомасштабним впровадженням інформаційних технологій в усі сфери людської діяльності, що, у свою чергу, також зумовлено процесами глобалізації світової спільноти. Характерним явищем глобалізації є великий обсяг інформації у текстовому представленні, в якому, з різних причин, дані про автора можуть бути відсутні. Тому здійснення авторської атрибуції тексту має велику вагу особливо у сфері судочинства (анонімні листи, заповіти та інші судові документи); у сфері науки (встановлення відсотку авторського внеску публікацій, що мають більше ніж одного автора); в офіційно-діловій сфері (документи); у сфері літератури (художні твори) та в інших сферах.

Встановлення авторства тексту здебільшого ґрунтується на визначенні частоти вживання мовних одиниць, яка є певним еталоном-індикатором авторського стилю. Відповідно до існуючих інформаційних систем, які використовуються у цій сфері, найчастіше дослідники вибирають рівень значущості 0,05 і тексти різних авторів відрізняються з довірчою ймовірністю 0,95. Однак у багатьох важливих випадках виникає необхідність підвищити довірчу ймовірність автоматизованої диференціації текстів. Зокрема, така потреба виникає при роботі із низкою офіційних документів. Власне вирішенню цього важливого завдання і присвячене дане дисертаційне дослідження.

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

Дисертаційна робота тісно пов'язана з планами науково-дослідної та навчальної роботи кафедри прикладної лінгвістики Національного університету “Львівська політехніка” – “Застосування формальних моделей для дослідження інформаційних природномовних процесів”. Дослідження, представлене в дисертації, проводилося у рамках держбюджетних науково-дослідних робіт Національного університету “Львівська політехніка”, а саме:

1. “Застосування формальних моделей для дослідження інформаційних природномовних процесів” (номер державної реєстрації 0114U001239).

2. ДБ/Нейрозахист “Нейромережева технологія захисту та передачі даних у реальному часі з використанням шумоподібних кодів” (номер державної реєстрації 0119U002256, січень 2019 по грудень 2020 рр.).

***Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих у дисертаційній роботі.*** Розроблені методи та вдосконалені моделі є науково і практично обґрунтованими. Обґрунтованість одержаних результатів дисертаційної роботи полягає у коректному використанні класичних методів математичної статистики. Припущення, покладені в основу теоретичних досліджень, є коректні і не суперечать особливостям предметної області. Достовірність наукових положень, висновків сформульованих у дисертаційній роботі підтверджуються низкою експериментальних досліджень програмних засобів автоматизації диференціації фоностатистичних структур стилів англійської мови та апробацією результатів на 8 наукових конференціях.

***Наукова новизна отриманих результатів.*** Основними результатами, особисто одержаними здобувачкою є:

1) вперше, розроблено метод комплексної диференціації фоностатистичних структур стилів англійської мови, який ґрунтується на поєднанні статистичних методів перевірки гіпотез, ранжування та визначення відстаней між стилями, що уможливило підвищення достовірності автоматизованої диференціації стилів із використанням сучасних програмних засобів;

2) вперше розроблено багатофакторний метод автоматизованого визначення ступеню дії факторів стилю, підстилю та авторської манери викладу, який ґрунтується на зіставленні фоностатистичних структур різних авторів, різних підстилів і різних стилів та забезпечує можливість підвищення достовірності авторської атрибуції тексту за його загальною стилевою маркованістю;

3) вдосконалено статистичний метод стилевої, підстилевої та авторської диференціації текстів з використанням методу перевірки статистичних гіпотез та методу ранжування, який відрізняється від існуючих розмежуванням текстів за кількістю груп приголосних фонем, що уможливорює підвищення достовірності оцінювання диференційних ознак стилів, підстилів та текстів різних авторів;

4) вдосконалено статистичний метод визначення стилерозрізняльної здатності груп приголосних фонем, який на відміну від інших, дає змогу визначити групу приголосних фонем з найвищою стилерозрізняльною здатністю і, тим самим, зменшити кількість груп приголосних фонем, за якими розмежовуються стилі, що підвищує достовірність та рівень автоматизації диференціації фоностатистичних структур стилів англійської мови.

### ***Практичні результати роботи, їх рівень та ступінь впровадження.***

Практичне значення одержаних результатів полягає у наступному:

1. Розроблено структуру програмної системи автоматизованої диференціації фоностатистичних структур стилів англійської мови, яка ґрунтується на модульному принципі, що дає змогу швидко модифікувати і вдосконалювати програмну систему.

2. Розроблено алгоритми реалізації методів та моделей диференціації фоностатистичних структур стилів англійської мови для ефективнішого розмежування стилів на фонологічному рівні.

3. Розроблено програмні засоби системи автоматизованої диференціації фоностатистичних структур стилів на мові програмування Java, що забезпечує кросплатформність програмного засобу та дає змогу автоматизувати процес авторської атрибуції тексту.

4. Отримані результати виконаного дослідження використано при розробці методичного забезпечення для дисциплін “Основи прикладної лінгвістики” і “Контрастивна стилістика англійської та української мов” у



навчальному процесі кафедри прикладної лінгвістики Національного університету “Львівська політехніка”, що підтверджено відповідним актом.

**Повнота викладення результатів досліджень в опублікованих працях.** Аналіз наукових результатів дисертаційної роботи Хомицької І.Ю. дає змогу зробити висновок про особистий внесок здобувачки у науку, що передбачає розроблення нових методів і програмних засобів автоматизації диференціації фоностатистичних структур стилів англійської мови.

Основні положення та результати роботи повністю викладено у 23 наукових працях, з яких: 8 статей у наукових фахових виданнях України, 5 з яких включені до міжнародної наукометричної бази Index Copernicus, 2 статті у науковому періодичному виданні України, що входить до міжнародної наукометричної бази Scopus, 1 стаття у науковому періодичному виданні Швейцарії, що входить до міжнародної наукометричної бази Scopus, та 12 публікацій у матеріалах конференцій, 11 з яких входять до міжнародних наукометричних баз Scopus та Web of Science.

**Оцінка основного змісту дисертації та її структури.** Дисертаційна робота є завершеною науковою роботою, стиль викладу думки характеризується виразністю і чіткістю, науково-технічна література використовується коректно. Дисертація складається з анотації, вступу, чотирьох розділів, висновків, списку використаної літератури та додатків.

У *вступі* обґрунтовано актуальність теми дисертаційної роботи, сформульовано мету та основні завдання дослідження, зазначено наукову новизну роботи, розглянуто практичне значення отриманих результатів і зв'язок роботи з науковими темами. Крім того, наведено інформацію про апробацію результатів роботи, особистий внесок автора та публікації.

У *першому розділі*, проведено аналіз існуючих методів, моделей та програмних засобів диференціації стилів та авторської атрибуції текстів.

За результатами проведеного аналізу використовуваних методів і моделей авторської атрибуції можна зробити висновок, що підвищення достовірності встановлення авторства тексту є актуальним науковим завданням, для розв'язання якого необхідно розробити нові та вдосконалити існуючі методи та моделі. У цьому контексті зазначено, що розроблені методи повинні базуватися на поєднанні статистичних критеріїв та враховувати

ступінь дії факторів стилю, підстилю та авторської манери викладу. У плані вдосконалення статистичних моделей необхідно спиратися на принцип розмежування авторських стилів за кількістю груп приголосних фонем та за стилерозрізняльною здатністю груп приголосних фонем. Розроблені методи та вдосконалені моделі повинні реалізуватися на мові програмування Java, яка є кросплатформною.

У *другому розділі*, представлено розроблені методи диференціації функціональних стилів англійської мови, зокрема: метод комплексного аналізу диференціації фоностатистичних структур стилів та багатофакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу.

У роботі розроблено метод комплексного аналізу диференціації фоностатистичних структур стилів, який поєднує такі статистичні методи: метод перевірки статистичних гіпотез (критерії Стюдента, Колмогорова-Смірнова та  $\chi^2$ -квадрат), метод ранжування та метод визначення відстаней між стилями. Розроблений метод забезпечує підвищення достовірності диференціації стилів застосуванням поєднання п'яти зазначених методів. За рівень значущості у ньому вибрано значення 0,05.

Другим розробленим методом є багатофакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу, який уможливорює підвищення достовірності диференціації авторських стилів за різницею значень загальної стилевої маркованості зіставлених текстів різних авторів (за запропонованою формулою).

У *третьому розділі*, описано побудовану модель системи фоностатистичних структур досліджуваних стилів (художнього, розмовного, газетного, публіцистичного, наукового) англійської мови, яка є аналогом системи функціональних стилів англійської мови. Складовими моделі системи фоностатистичних структур досліджуваних стилів є вдосконалені статистичні моделі: модель стилевої, підстилевої та авторської диференціації текстів за методом гіпотез і ранжування та модель визначення стилерозрізняльної здатності груп приголосних фонем досліджуваних стилів.

У *четвертому розділі*, описано розроблену програмну систему автоматизованої диференціації фоностатистичних структур функціональних стилів англійської мови та наведено результати дослідження.

Розроблено структуру та алгоритм програмної системи диференціації фоностатистичних структур функціональних стилів англійської мови. Модульний принцип побудови структури системи уможливило здійснення модифікації програмного забезпечення (ПЗ) та підвищення рівня автоматизації авторської атрибуції тексту. Систему розроблено на об'єктно-орієнтованій мові програмування Java.

*Висновки* до дисертаційного дослідження сформульовані чітко і повною мірою висвітлюють отримані в роботі результати.

*Список використаної літератури* достатньою мірою відображає предметну область дослідження та містить достатню кількість іноземних джерел.

**Відповідність дисертації та автореферату встановленим вимогам.** За своєю структурою, обсягом і оформленням дисертація та автореферат відповідають вимогам, встановленим до кандидатських дисертацій.

Основні положення автореферату та дисертації є ідентичними. Автореферат відображає інформацію, яка міститься у самій роботі.

Стиль викладу матеріалів досліджень, наукових положень і рекомендацій є доступний для розуміння.

Дисертація та автореферат мають завершену логічну структуру. Манера написання коректна, послідовна та грамотна. Авторка досягла поставленої мети та розв'язала сформульовані завдання. Основний зміст роботи відображено у висновках повною мірою.

*До недоліків та зауважень дисертаційної роботи можна віднести:*

1. В дисертації чомусь без належного обґрунтування вибрано класичні теоретико-ймовірнісні моделі і статистичні методи аналізу та диференціації текстових стилей. Однак видається більш доцільним для вирішення завдання дисертації застосувати сучасні методи машинного навчання, зокрема, нейромережеві технології, які мають ряд важливих переваг над класичними статистичними методами аналізу та перевірки гіпотез.
2. Було б доцільним в авторефераті та дисертації для оцінювання ефективності розробленого програмного забезпечення дослідити часову



складність розроблених алгоритмів та порівняти її із часовою складністю відомих методів.

3. У третьому розділі дисертації авторка не зовсім коректно використовує термін «модель» в таких конструктах як «статистична модель стилевої, підстилевої та авторської диференціації» та «статистична модель визначення стилерозрізняльної здатності груп приголосних фонем», оскільки більш коректним у цих випадках є використання терміну «метод», а саме – «статистичний метод стилевої, підстилевої та авторської диференціації» та «статистичний метод визначення стилерозрізняльної здатності груп приголосних фонем».
4. Не цілком зрозумілим є назва пропонованого авторкою методу, а саме «методу комплексного аналізу диференціації фоностатистичних структур», адже як можна аналізувати диференціацію? Можливо б вартувало цей метод назвати комплексним методом диференціації фоно-статистичних структур?
5. У результатах роботи, які стосуються диференціації авторської манери викладу у колективних статтях, варто б було навести специфічні ознаки, за якими можна було б розрізняти авторські стилі співавторів єдиного документу.
6. При викладенні матеріалу дисертації часто на передній план впливає стиль та питання філологічного характеру, що не характерно для дисертаційних досліджень за технічними науками.
7. З практичної точки зору дисертаційну роботу суттєво б підсилило наявність впровадження отриманих результатів не лише у навчальний процес організації, де виконана робота, але і у ряд інших підприємств та організацій, які мають стосунок до тематики дисертаційного дослідження.
8. У роботі подекуди зустрічаються граматичні помилки та описки.

Проте вказані зауваження не знижують загальної позитивної оцінки дисертаційної роботи.

### *Загальний висновок*

На підставі вивчення результатів дослідження наведених у дисертації, авторефераті та наукових працях здобувачки, опублікованих за темою дисертації, вважаю, що дисертаційна робота Хомицької І.Ю. є кваліфікаційною науковою працею, що містить нові науково-прикладні результати, які у своїй сукупності вирішують актуальне наукове завдання підвищення достовірності автоматизованої диференціації фоностатистичних структур стилів англійської мови на основі розроблених методів та програмних засобів. Тема, мета та поставлені завдання дисертаційної роботи відповідають науковій спеціальності 10.02.21 – структурна, прикладна і математична лінгвістика.

Автореферат відповідає змісту дисертації повною мірою та відображає суть отриманих результатів і висновків у дисертаційній роботі. Автореферат і дисертація оформлені згідно з вимогами до кандидатських дисертацій.

З огляду на актуальність теми дисертації, наукової і практичної цінності отриманих результатів дослідження та врахування у дисертації вимог пп. 9, 11, 12 «Порядку присудження наукових ступенів», вважаю, що її авторка Хомицька Ірина Юріївна заслуговує присудження їй наукового ступеня кандидата технічних наук за спеціальністю 10.02.21 – структурна, прикладна і математична лінгвістика.

Офіційний опонент

професор кафедри комп'ютерних систем та мереж  
Тернопільського національного технічного  
університету імені Івана Пулюя, МОН України,  
доктор технічних наук, професор

С.А. Лупенко

Підпис професора Лупенка С.А. засвідчую  
Проректор з наукової роботи  
Тернопільського національного технічного  
університету імені Івана Пулюя, МОН України,  
доктор технічних наук, професор



П.О. Марушак