

Міністерство освіти і науки України
Національний університет “Львівська політехніка”

ХОМИЦЬКА ІРИНА ЮРІЇВНА



УДК 519.76:811.111'32(043.3)

Методи та засоби диференціації фоностатистичних структур функціональних стилів англійської мови

Спеціальність 10.02.21 – структурна, прикладна та математична лінгвістика

АВТОРЕФЕРАТ

дисертації на здобуття наукового ступеня

кандидата технічних наук

Львів – 2021

Дисертацією є рукопис.

Робота виконана у Національному університеті “Львівська політехніка” Міністерства освіти і науки України.

Науковий керівник:

доктор технічних наук, професор
ТЕСЛЮК Василь Миколайович,
Національний університет «Львівська політехніка»,
в.о. зав. кафедри автоматизованих систем управління.

Офіційні опоненти:

доктор технічних наук, професор
ЛУПЕНКО Сергій Анатолієвич,
Тернопільський національний технічний
університет імені Івана Пулюя,
професор кафедри комп'ютерних
систем та мереж

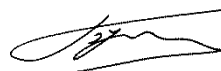
кандидат технічних наук,
ОСТАПОВА Ірина Вадимівна,
Український мовно-інформаційний фонд
НАН України,
старший науковий співробітник

Захист відбудеться 28 квітня 2021 р. о 15⁰⁰ годині на засіданні спеціалізованої вченої ради Д 35.052.05 у Національному університеті “Львівська політехніка” (79013, м. Львів, вул. С. Бандери, 12, ауд. 226 головного корпусу).

З дисертацією можна ознайомитися у науково-технічній бібліотеці Національного університету “Львівська політехніка” (79013, м. Львів, вул. Професорська, 1).

Автореферат розісланий « 19 » березня 2021 р.

Учений секретар
спеціалізованої вченої ради,
доктор технічних наук, професор



Р. А. Бунь

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Широкомасштабне впровадження інформаційних технологій в усі сфери людської діяльності забезпечує глобалізацію світової спільноти. Глобалізація передбачає опрацювання великого обсягу інформації у текстовому представленні, в якому, з різних причин, дані про автора можуть бути відсутні. У цьому випадку виникає завдання авторської атрибуції тексту. Встановлення авторства тексту є завданням великої ваги у судочинстві (анонімні листи, заповіти та інші судові документи); у сфері науки (встановлення відсотку авторського внеску статей, що мають більше ніж одного автора); в офіційно-діловій сфері (документи); у сфері літератури (художні твори) та в інших сферах.

Встановлення авторства тексту ґрунтується на визначенні частоти вживання мовних одиниць, яка є диференційною характеристикою авторського стилю. Аналіз сучасного стану досліджень, в яких застосовуються методи математичної статистики, показав, що значний внесок у розроблення методів, моделей та засобів диференціації текстів зроблено: В. Перебийніс – визначено статистичні параметри стилів української мови; В. Левицьким – визначено, які статистичні методи є ефективними для фонологічного рівня мови; Д. Сегалом – сформульовано основи фонологічної статистики; Ш. Аргамомом, М. Копелем, Дж. Шлером – розроблено методика встановлення авторства анонімних листів та ін.

Зазначивши вагомий доробок попередніх досліджень, слід наголосити, що застосовані дослідниками методи дали змогу отримати різні значення достовірності авторської атрибуції тексту (ААТ). Відповідно до існуючих результатів, найчастіше дослідники вибирають рівень значущості 0,05, який є оптимальним, і тексти різних авторів відрізняються з довірчою ймовірністю 0,95. Оскільки низка офіційних документів потребує високого рівня достовірності, його можна отримати застосувавши поєднання ефективних статистичних методів, які дають однаковий результат на тому самому матеріалі. Тому з погляду сучасного стану досліджень, розроблення методів для підвищення достовірності диференціації текстів є актуальним науковим завданням. Таким чином, тема дисертаційного дослідження “Методи та засоби диференціації фоностатистичних структур функціональних стилів англійської мови” є актуальною.

Зв’язок роботи з науковими програмами, планами, темами. Тема дисертації відповідає науковому напряму кафедри прикладної лінгвістики Національного університету “Львівська політехніка” – “Застосування формальних моделей для дослідження інформаційних природномовних процесів”. Дисертаційна робота виконана у межах держбюджетних науково-дослідних робіт – “Застосування формальних моделей для дослідження інформаційних природномовних процесів” (номер державної реєстрації 0114U001239) та 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.).

Мета і завдання дослідження. Метою дослідження є підвищення достовірності диференціації функціональних стилів англійської мови (розмовного, художнього, газетного, публіцистичного, наукового) на основі розроблених

методів, моделей та програмних засобів. Для досягнення поставленої мети необхідно розв'язати такі завдання:

- провести аналіз використовуваних методів, моделей та програмних засобів диференціації стилів і текстів різних авторів;
- розробити метод комплексного аналізу диференціації фоностатистичних структур стилів для підвищення достовірності стилевої диференціації текстів;
- розробити багатофакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу, який дає змогу підвищити достовірність здійснення авторської атрибуції тексту за його загальною стилевою маркованістю;
- вдосконалити статистичну модель стилевої, підстилевої та авторської диференціації за методом гіпотез і методом ранжування, яка робить можливим визначення диференційних ознак стилів, підстилів та текстів різних авторів з більшою достовірністю;
- вдосконалити статистичну модель визначення стилерозрізняльної здатності груп приголосних фонем, яка забезпечує зменшення кількості груп приголосних фонем та підвищення рівня автоматизації диференціації текстів за групою приголосних фонем з найвищою стилерозрізняльною здатністю;
- розробити програмні засоби для автоматизації диференціації стилів англійської мови (художнього, розмовного, наукового, газетного, публіцистичного).

Об'єктом дослідження є процес диференціації фоностатистичних структур функціональних стилів англійської мови (художнього, розмовного, наукового, газетного, публіцистичного).

Предметом дослідження є математичні методи, моделі та програмні засоби диференціації фоностатистичних структур досліджуваних стилів англійської мови.

Методи дослідження. Застосовано статистичні методи – метод гіпотез, метод ранжування та метод визначення відстаней між стилями для диференціації фоностатистичних структур стилів англійської мови. Використано методи статистичного моделювання для визначення взаємодії фоностатистичних структур стилів англійської мови. Використано методи теорії алгоритмів та методи об'єктно-орієнтованого підходу в процесі розроблення програмних засобів.

Наукова новизна одержаних результатів. Новизною дослідження є розроблені методи та вдосконалені статистичні моделі диференціації стилів англійської мови:

1) вперше розроблено метод комплексного аналізу диференціації фоностатистичних структур стилів англійської мови, який ґрунтується на поєднанні статистичних методів гіпотез, ранжування і визначення відстаней між стилями та дає змогу підвищити достовірність диференціації стилів;

2) вперше розроблено багатофакторний метод 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 статистичну модель стилевої, підстилевої та авторської диференціації за методом гіпотез та методом ранжування, яка відрізняється від

існуючих розмежуванням текстів за кількістю груп приголосних фонем, що уможлиблює підвищення достовірності визначення диференційних ознак стилів, підстилів та текстів різних авторів;

4) вдосконалено статистичну модель визначення стилерозрізняльної здатності груп приголосних фонем, яка 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 текстів.

Практичне значення одержаних результатів полягає у наступному:

1. Розроблено структуру системи 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 як методичне 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.

Особистий внесок здобувача. Всі наукові результати теоретичних і практичних досліджень, викладених у дисертації, одержано автором особисто. У працях, опублікованих у співавторстві, здобувачеві належать: метод комплексного аналізу диференціації фоностатистичних структур стилів англійської мови та алгоритм його реалізації [5]; статистична модель стилевої диференціації наукового, розмовного, газетного і художнього стилів за методом гіпотез, результати досліджень [4, 9, 14]; багатofакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу та алгоритм його реалізації [3, 22]; статистична модель визначення стилерозрізняльної, підстилерозрізняльної та авторорозрізняльної здатності груп приголосних фонем, результати досліджень [1, 6, 8, 10, 11, 13]; статистична модель авторської диференціації за методом гіпотез і ранжування, результати досліджень [7, 18]; статистична модель підстилевої диференціації за методом гіпотез, результати досліджень [12, 16]; статистична модель стилевої диференціації за групою середньоязикових фонем за методом гіпотез, результати досліджень [15]; статистична модель стилевої диференціації за методами гіпотез, ранжування і визначення відстаней між стилями, результати досліджень [17]; алгоритм, структура та програмне забезпечення системи диференціації фоностатистичних структур стилів, результати досліджень [19]; алгоритм визначення стилерозрізняльної здатності груп приголосних фонем, результати досліджень [2, 20, 21]; алгоритм ААТ, результати досліджень [23].

Апробація роботи. Викладені у дисертації результати досліджень представлено на 7 міжнародних науково-технічних конференціях: XIth, XIIth, XIIIth, and XIVth Scientific and Technical Conferences on Computer Science and Information Technologies (CSIT'2016, CSIT'2017, CSIT'2018, CSIT'2019), Lviv, 2016, 2017, 2018, 2019; XIVth and XVth Scientific and Technical Conferences on The Experience of Designing and Application of Computer Aided Design Systems (CADMS'2017, CADMS'2019): Polyana, 2017, 2019; 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020), Lviv, 2020.

Публікації. 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.

Структура та обсяг роботи. Дисертація складається з анотації, вступу, чотирьох розділів, висновків, списку використаної літератури та додатків. Загальний обсяг дисертації становить 262 сторінки, у тому числі 132 сторінки основного тексту, 34 рисунки та 98 таблиць, список використаної літератури налічує 184 бібліографічних найменувань. Дисертація містить 9 додатків, розміщених на 91 сторінці.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність теми дисертаційної роботи, сформульовано мету та основні завдання дослідження, визначено наукову новизну роботи, розглянуто практичне значення отриманих результатів і зв'язок роботи з науковими темами. 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.

Застосування ймовірнісних методів та методів математичної статистики є доречним при дослідженні явищ мови, яка є ймовірнісною системою (роботи Г. Альтмана, В. Перибийніс, В. Левицького). Аналіз робіт зазначених дослідників, в яких стилі диференціюються на фонологічному рівні, дає змогу ствердити, що рівень достовірності не є достатнім, бо тексти не розмежовуються потужним непараметричним критерієм Колмогорова-Смірнова у поєднанні з іншими критеріями. Крім того, зважаючи на те, що різні статистичні методи дають різні результати при зіставленні тих самих двох текстів, застосування лише одного статистичного методу, не може дати вичерпної інформації про достовірність отриманих результатів. У цьому контексті, логічним є пошук поєднання найдієвіших статистичних методів. Поєднання потужних параметричних і непараметричних критеріїв диференціації стилів забезпечує підвищення рівня достовірності результатів. Оскільки ідея поєднання використовуваних статистичних методів є відсутньою в сучасних дослідженнях, актуальним є

розроблення методу комплексного аналізу диференціації стилів. Крім вибору оптимального методу диференціації текстів, не менш важливим є вибір рівня мови, на якому диференціюються стилі. Рівень мови визначає достовірність отриманих результатів. Аналізуючи рівні мови з погляду ступеню впорядкованості елементів та змінності їх кількості, оптимальним рівнем можна вважати фонологічний, бо кількість його елементів є незмінною, його структура є строгою і його найлегше формалізувати. Беручи до уваги зазначені аргументи, можна зробити висновок, що для підвищення достовірності диференціації стилів, актуальним науковим завданням є розроблення методу комплексного аналізу диференціації фоностатистичних (статистичних структур на фонологічному рівні) структур стилів.

Проведений аналіз використовуваних методів показав, що в сучасних дослідженнях відсутній багатофакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу. Ідея методу полягає у визначенні загальної стилевої маркованості досліджуваного тексту, яка є рівнодіючою факторів стилю, підстилю та авторської манери викладу. Різниця значень загальної стилевої маркованості досліджуваного тексту, за якою визначаються істотні відмінності між авторськими стилями, дає змогу підвищити достовірність авторської диференціації текстів. Тому, розроблення багатофакторного методу є актуальним завданням для досягнення мети дисертаційного дослідження.

На основі проведеного аналізу використовуваних моделей стилевої диференціації стилів, слід зазначити вагомий доробок у розроблення статистичних моделей стилевої диференціації за допомогою параметричного критерію (критерію Стьюдента) у роботах В. Перибийніс. Однак, для підвищення достовірності стилевої диференціації, слід наголосити на необхідності вдосконалення існуючих статистичних моделей, застосовуючи потужний непараметричний критерій Колмогорова-Смірнова та диференціюючи стилі за кількістю груп приголосних фонем, за якими встановлено істотні відмінності за методом гіпотез та ранжування. Для підвищення рівня достовірності та автоматизації стилевої, підстилевої та авторської атрибуції тексту, необхідно вдосконалити статистичну модель визначення стилерозрізняльної здатності груп приголосних фонем досліджуваних стилів на основі визначення групи фонем з найбільшою стилерозрізняльною здатністю та диференціювати стилі за цією групою фонем.

Аналіз засобів диференціації текстів показав, що ефективними є програмні системи від NeoNeuro, Quora, JGAAP, JGAAP in version 3.1. [<https://neoneuro.com/ru/products/authorship-attribution>], [<https://www.quora.com/What-is-the-best-tool-or-software-for-stylometry>], [<http://www.evllabs.com>], [<http://www.jgaap.com>]. Однак програми опрацьовують тексти на лексико-семантичному рівні, структура якого є менш строгою ніж фонологічного рівня (транскрипційні символи), і його важче формалізувати. Тому достовірність результатів на лексико-семантичному рівні є нижчою, ніж на фонологічному рівні. Зважаючи на це, актуальним є розроблення програмної системи авторської, підстилевої і стилевої атрибуції тексту на фонологічному рівні.

У другому розділі розроблено методи диференціації функціональних стилів англійської мови, зокрема: *метод комплексного аналізу диференціації*

фоностатистичних структур стилів та багатofакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу.

Для підвищення достовірності розмежування стилів, підстилів та текстів різних авторів, розроблено метод комплексного аналізу диференціації фоностатистичних структур стилів, який поєднує такі статистичні методи: метод гіпотез (критерії Стьюдента, Колмогорова-Смірнова та хі-квадрат), метод ранжування та метод визначення відстаней між стилями. Підвищення достовірності диференціації стилів забезпечується застосуванням п'яти методів, одним з яких є непараметричний і його потужність прямує до 1 (критерій Колмогорова-Смірнова). Результати отримано з довірчою ймовірністю 0,95 (рис. 1).

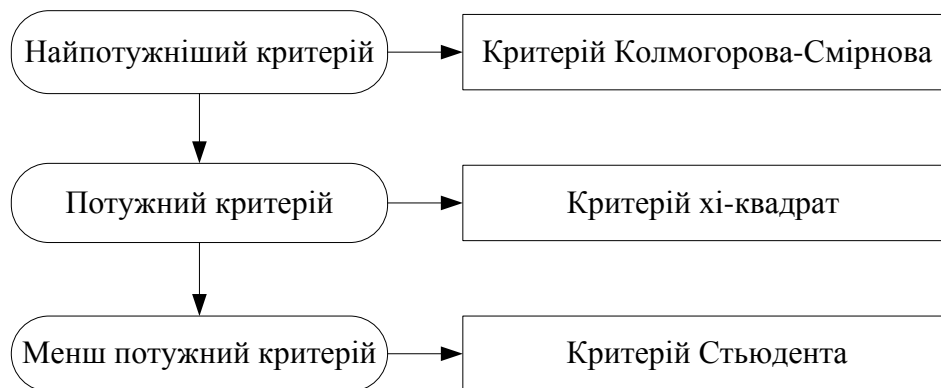


Рисунок 1 – Структурна схема рівня потужності критеріїв диференціації стилів

Розроблений метод реалізується наведеним алгоритмом, а саме:

Крок 1. Перетворити текст англійською мовою в його транскрипційний варіант.

Крок 2. Сформувати вибірки приголосних фонем обсягом в 51000 фонем.

Крок 3. Поділити вибірки на 51 порцію.

Крок 4. Об'єднати приголосні фонем в групи.

Крок 5. Обчислити абсолютну та середню частоти по порціях і групах:

$$\bar{X} = \sum_{i=1}^m x_i n_i / N$$
, де \bar{X} – середня частота вживання фонем; n_i – кількість частот, які потрапляють в i -й інтервал; x_i – середина інтервалу; m – кількість інтервалів; N – кількість порцій вибірки.

Крок 6. Побудувати спадні ряди середніх частот для кожної групи фонем.

Крок 7. Визначити істотні відмінності між попарно зіставленими текстами за різницею рангових показників.

Крок 8. Вибрати за критерій диференціації стилів – середню частоту груп приголосних фонем.

Крок 9. Перевірити відповідність частот груп приголосних фонем закону нормального розподілу (за критерієм Пірсона та спрощеним критерієм Романовського):

$$x^2 = \sum_{i=1}^k \frac{(n_i - n'_i)^2}{n'_i}; \frac{|x^2 - \nu|}{\sqrt{2\nu}} \geq 3$$
, де i – порядковий номер інтервалу; k – кількість інтервалів; n_i – емпірична частота вживання фонем; n'_i – теоретична частота вживання фонем; x^2 – значення статистики критерію Пірсона; ν – значення кількості ступенів вільності.

Крок 10. Розмежувати тексти за критерієм Стьюдента.

Крок 11. Визначити відстані між стилями за формулою: $l = \frac{t - t_0}{t}$, де t – значення статистики Стьюдента; t_0 – табульована величина.

Крок 12. Розмежувати тексти за критерієм Колмогорова-Смірнова.

Крок 13. Розмежувати тексти за критерієм хі-квадрат.

Крок 14. Визначити кількість груп приголосних фонем, за якими встановлено істотні відмінності у попарному зіставленні текстів.

Наведений алгоритм розробленого методу уможливив отримання результатів з довірчою ймовірністю 0,95. Достовірність диференціації стилів, підстилів та текстів різних авторів забезпечується застосуванням розглянутого поєднання методів, яке включає непараметричний метод Колмогорова-Смірнова з потужністю, що прямує до 1.

Для підвищення достовірності встановлення стилевої, підстилевої та авторської приналежності досліджуваного тексту, розроблено багатофакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу, який передбачає визначення загальної стилевої маркованості досліджуваного тексту (за запропонованою формулою). Особливістю методу є розмежування текстів за значенням загальної стилевої маркованості зіставлюваних текстів. Розроблений метод реалізується за схемою стиль→підстиль→автор. Алгоритм застосування методу містить такі основні кроки:

Крок 1. Визначити за критеріями Стьюдента, Колмогорова-Смірнова та хі-квадрат істотні відмінності у попарному зіставленні текстів: різних стилів, різних підстилів та різних авторів.

Крок 2. Визначити ступінь дії фактора стилю t_{f_1} за фіксованою групою приголосних фонем a : $\bar{x}_{1c}^a - \bar{x}_{2c}^a$ (c – стиль).

Крок 3. Визначити ступінь дії фактора підстилю t_{f_2} : $\bar{x}_{1n}^a - \bar{x}_{2n}^a$ (n – підстиль).

Крок 4. Визначити ступінь дії фактора авторської манери викладу t_{f_3} : $\bar{x}_{1a}^a - \bar{x}_{2a}^a$ (a – авторський стиль).

Крок 5. Визначити загальну стилеву маркованість досліджуваного тексту за формулою: $sm = t_{f_1} + t_{f_2} + t_{f_3} / 3$.

Крок 6. Визначити відмінності між авторськими стилями за різницею значень загальної стилевої маркованості: $sm_1 - sm_2$, де sm_1 – значення загальної стилевої маркованості тексту першого автора, а sm_2 – значення загальної стилевої маркованості тексту другого автора.

Отже, розроблений метод, враховуючи визначені значення дії трьох факторів (стилю, підстилю та авторської манери викладу), дає змогу з більшою

достовірністю здійснити стилеву, підстилеву та авторську атрибуцію досліджуваного тексту за різницею значень загальної стилевої маркованості зіставлюваних текстів з довірчою ймовірністю 0,95.

У третьому розділі побудовано 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.*

Для розв'язання завдання підвищення достовірності визначення диференційних ознак стилів, підстилів та текстів різних авторів, вдосконалено статистичну модель стилевої, підстилевої та авторської диференціації текстів за методами гіпотез і ранжування, яка на відміну від існуючих, враховує позицію фонему у слові, визначає відношення кількості груп приголосних фонем, за якими встановлено істотні відмінності, між попарно зіставленими стилями до загальної кількості груп приголосних фонем. Ідея розмежування стилів за кількістю груп приголосних фонем пропонується вперше. Велика кількість груп приголосних фонем (6 – 8), за якими тексти відрізняються істотно, дає змогу з більшою достовірністю робити висновки про їх відмінність.

Вдосконалена статистична модель стилевої, підстилевої та авторської диференціації текстів за методом гіпотез і ранжування реалізується наступним алгоритмом:

Крок 1. Визначити істотні відмінності для 8 груп приголосних фонем з довірчою ймовірністю 0,95 у зіставленні стилів, підстилів і текстів різних авторів, перевіряючи статистичні гіпотези з використанням критеріїв Стюдента, Колмогорова-Смірнова та χ^2 -квадрат.

Крок 2. Розмежувати стилі, підстилі та тексти різних авторів за встановленими істотними відмінностями, використовуючи метод ранжування.

Крок 3. Підрахувати кількість груп приголосних фонем, за якими встановлено істотні відмінності.

Крок 4. Визначити ступінь встановлених істотних відмінностей (DSD) за відношенням кількості груп приголосних фонем, за якими встановлено істотні відмінності між попарно зіставленими стилями ($SDGN$) до загальної кількості груп приголосних фонем (TGN):

$$DSD = \frac{SDGN}{TGN}. \quad (1)$$

Крок 5. Встановити низький ступінь встановлених істотних відмінностей ($LDSN$) за відношенням 1 – 3 груп приголосних фонем, за якими встановлено істотні відмінності між попарно зіставленими стилями, до загальної кількості груп приголосних фонем – 8: $LDSN = \frac{1}{8} = 0,13$; $LDSN = \frac{2}{8} = 0,25$; $LDSN = \frac{3}{8} = 0,38$.

Крок 6. Визначити середній ступінь встановлених істотних відмінностей ($MDSN$) за відношенням 4 – 5 груп приголосних фонем, за якими попарно

зіставлені стилі відрізняються істотно, до загальної кількості груп приголосних фонем (8): $MDSN = \frac{4}{8} = 0,5$; $MDSN = \frac{5}{8} = 0,63$.

Крок 7. Встановити високий ступінь встановлених істотних відмінностей (*HDSN*) за відношенням 6 – 8 груп приголосних фонем, за якими визначено істотні відмінності у попарному зіставленні стилів, до 8 груп приголосних фонем: $HDSN = \frac{6}{8} = 0,75$, $HDSN = \frac{7}{8} = 0,88$, $HDSN = \frac{8}{8} = 1$.

Таким чином, вдосконалена статистична модель дає змогу встановити ступінь диференціації стилів за критеріями перевірки гіпотези на однорідність вибірок та за різницею рангових показників. Якщо кількість груп приголосних фонем, за якими встановлено істотні відмінності між попарно зіставленими стилями дорівнює загальній кількості груп приголосних фонем, то ступінь диференціації стилів є найвищим. Якщо кількість груп приголосних фонем, за якими встановлено істотні відмінності між попарно зіставленими стилями дорівнює 1, то ступінь диференціації стилів є найнижчим. Отже, встановлення ступеню диференціації стилів з довірчою ймовірністю 0,95, дає змогу підвищити достовірність результатів для стилевої, підстилевої та авторської атрибуції досліджуваного тексту за кількістю груп фонем, за якими вибірки відрізняються істотно.

Для розв'язання завдання підвищення рівня автоматизації диференціації стилів, вдосконалено статистичну модель визначення стилерозрізняльної здатності груп приголосних фонем, яка базується на визначенні відношення кількості порівнянь, за якими виявлено істотні відмінності між стилями – до кількості всіх порівнянь. Зменшення кількості груп приголосних фонем, за якими диференціюються стилі, спрощує процес атрибуції тексту, підвищує його достовірність і забезпечує вищий рівень автоматизації шляхом здійснення атрибуції тексту за групою приголосних фонем з найбільшою стилерозрізняльною здатністю. Вдосконалена статистична модель визначення стилерозрізняльної здатності груп приголосних фонем досліджуваних стилів реалізується наступним алгоритмом:

Крок 1. Розмежувати тексти за критерієм Стюдента, Колмогорова-Смірнова та критерієм χ^2 -квадрат.

Крок 2. Визначити групу фонем, за якою встановлено істотні відмінності за критерієм Стюдента, Колмогорова-Смірнова та критерієм χ^2 -квадрат:

$$ADC_1 = \frac{t + \lambda_{n,m} + \hat{\chi}_n^2}{3}, \quad (2)$$

де ADC_1 – здатність групи фонем розрізняти авторські стилі; t – значення критерію Стюдента, $\lambda_{n,m}$ – значення критерію Колмогорова-Смірнова, $\hat{\chi}_n^2$ – значення критерію χ^2 -квадрат.

Крок 3. Визначити групу фонем, за якою встановлено істотні відмінності у всіх випадках позиції фонем в слові:

$$ADC_2 = \frac{UP + IP + FP}{3}, \quad (3)$$

де ADC_2 – здатність групи фонем розрізняти авторські стилі у трьох випадках позиції фонем в слові: UP – невизначена позиція фонем в слові, IP – фонема на початку слова, FP – фонема в кінці слова.

Крок 4. Визначити авторорозрізняльну здатність групи приголосних фонем (ADC) за відношенням кількості зіставлень, у яких встановлено істотні відмінності між текстами різних авторів (SSD) до загальної кількості всіх зіставлень (TNC):

$$ADC = \frac{SSD}{TNC}. \quad (4)$$

Крок 5. Визначити групу фонем з найбільшим значенням ADC .

Крок 6. Розмежувати тексти за цією групою фонем.

Отже, вдосконалена статистична модель дає змогу встановити групу фонем з найвищою стилерозрізняльною здатністю, що зменшує кількість груп фонем і забезпечує вищий рівень ААТ.

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

Розроблено структуру програмної системи диференціації фоностатистичних структур функціональних стилів англійської мови (рис. 2), яка ґрунтується на модульному принципі, що дає змогу швидко модифікувати розроблене програмне забезпечення (ПЗ), підвищити рівень автоматизації та ефективність стилевої, підстилевої та ААТ.

Перший модуль структури системи призначено для введення-виведення даних вибірок стилів, обсягом в 51 тис. фонем кожна. Другий модуль призначено для перетворення англійських текстів стилів у транскрипційні варіанти. Третій модуль призначено для визначення кількості приголосних фонем у кожній із 14 вибірок. Четвертий модуль призначено для визначення середніх частот груп губних, передньоязикових, середньоязикових, задньоязикових, носових, сонорних, щілинних та зімкнених фонем. П'ятий модуль призначено для перевірки відповідності частот груп приголосних фонем закону нормального розподілу. Математичне забезпечення модуля ґрунтується на використанні критерію Пірсона. Шостий модуль призначено для диференціації досліджуваних стилів. Математичне забезпечення модуля базується на використанні критерію Стюдента, критерію Колмогорова-Смірнова та критерію хі-квадрат.

Розроблений алгоритм функціонування програми автоматизації диференціації фоностатистичних структур функціональних стилів англійської мови реалізує такі кроки:

Крок 1. Вибрати тексти із п'яти стилів і трьох підстилів та завантажити у файлах з розширенням .txt.

Крок 2. Створити транскрипційні варіанти для всіх чотирнадцяти текстів стилів та підстилів.

Крок 3. Сформувані вибірки заданого обсягу (51 тис. приголосних фонем).

Крок 4. Визначити частоти приголосних фонем у чотирнадцяти вибірках.

Крок 5. Поділити вибірки, кожного із п'яти стилів та трьох підстилів, на 51 порцію.

Крок 6. Визначити значення частоти та середнє значення частоти приголосних фонем по порціях для художнього, газетного, публіцистичного, розмовного і наукового стилів.

Крок 7. Об'єднати фонем у вісім груп за акустико-артикуляційними ознаками.

Крок 8. Визначити середнє значення групи приголосних фонем у кожному стилі та підстилі.

Крок 9. Визначити теоретичний нормальний розподіл на основі емпіричного нормального розподілу.

Крок 10. Визначити теоретичну частоту.

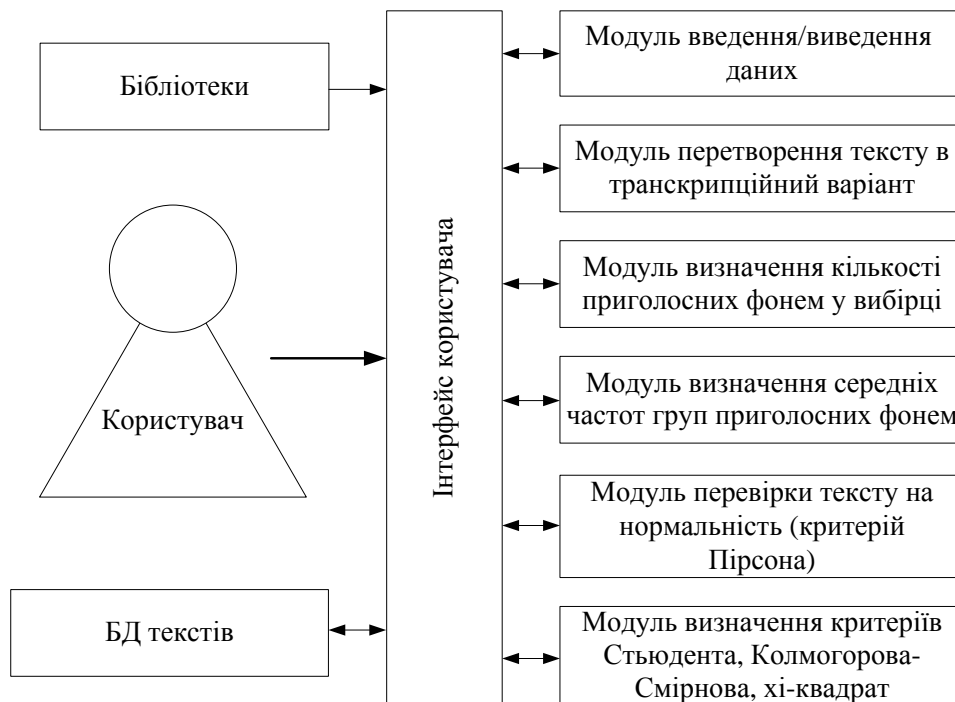


Рисунок 2 – Структура програмної системи диференціації стилів

Крок 11. Перевірити відповідність частот груп приголосних фонем закону нормального розподілу за критерієм згоди Пірсона.

Крок 12. Розмежувати дві зіставлені попарно вибірки за фіксованою групою приголосних фонем за критеріями Стьюдента, Колмогорова-Смірнова та хі-вадрат.

Розроблений алгоритм уможливує здійснення стилевої, підстилевої та авторської атрибуції досліджуваного тексту на фонологічному рівні, який характеризується строгішою впорядкованістю елементів, і тому, отримані на цьому рівні результати мають більшу достовірність.

Розроблено програмне забезпечення системи диференціації фоностатистичних структур стилів англійської мови. Розроблене ПЗ є кросплатформним, оскільки розроблення системи здійснено на об'єктно-орієнтованій мові програмування Java. Побудована структура класів ПЗ системи диференціації стилів зображена на рис. 3.

У класі Main використовується пакет flow, який забезпечує завантаження

файлів, послідовність виконання операцій, отримання результатів опрацювання даних. Для формування вибірки з приголосних фонем створено клас ConsonantProcessor. Наступним кроком є обчислення кількості приголосних фонем у кожній порції і об'єднання їх у групи за допомогою створеного класу ConsonantUtils, використовуючи метод countConsonantTypes(), який викликається для кожної групи фонем. Клас StatisticProcessor визначає критерій, який потрібно застосовувати. У цьому класі використано метод process, який за параметри приймає назву критерію і всі вибірки, які попарно зіставляються. Приклад меню вибору критерію диференціації текстів зображено на рис. 4.



Рисунок 3 – Структура класів ПЗ системи диференціації фоностатистичних структур стилів

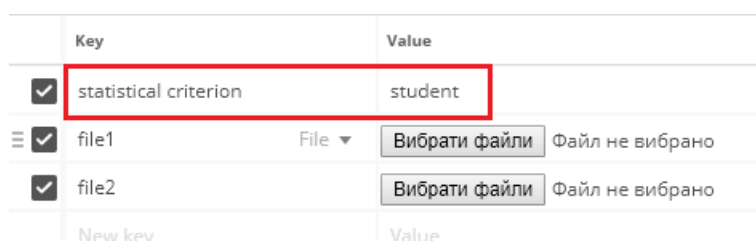


Рисунок 4 – Приклад меню вибору критерію розмежування досліджуваних текстів

Програмна реалізація розроблених методів і вдосконалених моделей дала змогу отримати нижченаведені результати.

У попарному зіставленні художнього, розмовного, газетного, публіцистичного та наукового стилів англійської мови, визначено різний ступінь відмінності порівнюваних текстів за різними групами приголосних фонем. Значні істотні відмінності отримано у зіставленні наукового (тексти з технічної фізики (Acta Physica Polonica)) і розмовного стилів (розмовно-побутова лексика (Free French Travel Phrasebook)) за групою передньоязикових фонем (табл. 1). Достовірність результатів забезпечено запропонованим поєднанням статистичних методів та використанням потужного непараметричного критерію Колмогорова-Смірнова.

Таблиця 1 – Результати зіставлення наукового і розмовного стилів

Зіставлені стилі за групою передньоязикових фонем	Метод гіпотез	Метод ранжування	Метод визначення відстаней між стилями
Науковий-Розмовний	$2Q < 0,1\%$	різниця рангових показників – 6	0,78

Авторську атрибуцію тексту здійснено розробленим багатофакторним методом визначення ступеню дії факторів стилю, підстилю та авторської манери викладу за групою передньоязикових фонем для трьох випадків позиції фонем в слові у зіставленні текстів поезії Байрона і Мура (табл. 2). Істотна відмінність

авторських стилів Байрона і Мура встановлена за різницею значень загальної стилевої маркованості зіставлених текстів поезії і дорівнює 1,8 (табл. 2). Отриманий результат підтверджує відмінність авторських стилів.

Таблиця 2 – Результати зіставлення поезії Байрона і Мура

Зіставлені тексти Байрона і Мура за групою передньоязикових фонем	Загальна стилева маркованість поезії Байрона	Загальна стилева маркованість поезії Мура	Значення істотної відмінності
Поезія Байрона-поезія Мура	13,3	11,5	1,8

Встановлений показник відмінності авторських стилів Байрона і Мура свідчить про дієвість багатофакторного методу, який забезпечує можливість підвищення достовірності ідентифікації автора поетичного твору у зіставленні творів різних поетів. Довірча ймовірність результатів – 0,95.

Відмінність та схожість лексико-семантичного наповнення зіставляваних текстів визначено реалізацією вдосконаленої статистичної моделі стилевої, підстилевої та авторської диференціації текстів за методом гіпотез (критерієм Стюдента), яка дала змогу встановити кількість груп приголосних фонем, за якими стилі відрізняються істотно. Диференціацію стилів за кількістю груп приголосних фонем здійснено вперше для підвищення достовірності диференціації стилів. У табл. 3 наведено результати для випадку не врахування позиції фонем в слові. Значні істотні відмінності встановлено за 6 – 7 групами фонем у зіставленні розмовного (РС) і наукового стилів (НС); середні істотні відмінності виявлено за 4 – 5 групами фонем – у порівнянні газетного стилю (ГС) і прози Байрона (Пр), газетного стилю (ГС) і драми Шоу (Др); незначні істотні відмінності встановлено за 1 – 3 групами фонем у зіставленні поезії Байрона (ПзБ) і Мура (ПзМ).

Таблиця 3 – Відмінності між стилями за кількістю груп приголосних фонем при не врахуванні позиції фонем у слові

Стиль	ПзБ	ПзМ	Пр	Др	ГС	НС	РС
	I	I	I	I	I	I	I
ПзБ	-	3	4	6	1	2	6
ПзМ	3	-	3	7	6	6	6
Пр	4	3	-	7	4	6	7
Др	6	7	7	-	6	6	4
ГС	1	6	4	6	-	3	6
НС	2	6	6	6	3	-	7
РС	6	6	7	4	6	7	-

Вдосконалена статистична модель авторської диференціації текстів забезпечила визначення відмінності авторських стилів у зіставлених вибірках з

публіцистичного і художнього стилів. Результати реалізації моделі наведено у табл. 4. За результатами тестування моделі встановлено, що найпотужнішим критерієм є критерій Колмогорова-Смірнова, за яким тексти відрізняються за всіма вісьмома групами приголосних фонем.

Встановлені істотні відмінності між текстами Б. Обама, Д. Трамп, Д. Вебстера, С. Логан, Е. Бронте і В. Текерея за трьома критеріями, за кількістю груп приголосних фонем, підтверджують своєрідність специфіки авторської манери викладу у текстах схожої тематики і уможливають підвищення достовірності диференціації інших текстів цих авторів.

Таблиця 4 – Результати визначення авторської диференціації за методом гіпотез

Зіставлені авторські стилі	Критерій Стьюдента	Критерій Колмогорова-Смірнова	Критерій χ^2 -квадрат
Обама-Трамп	-	8	6
Обама-Вебстер	-	8	6
Обама-Логан	-	8	6
Трамп-Вебстер	2	8	7
Трамп-Логан	2	8	6
Логан-Вебстер	3	8	6
Бронте-Текерей	2	8	3

Поєднання трьох різних методів: методу гіпотез (МГ), методу ранжування (МР) та методу визначення відстаней між стилями (МВС) дало змогу отримати середні істотні відмінності у зіставленні розмовного і газетного стилів (РС і ГС) за групою сонорних фонем для 3 випадків позиції фонем в слові, що зображено на рис. 5.

У даному зіставленні стилів кожен із застосованих трьох методів забезпечив отримання середніх істотних відмінностей (числові показники наведено на рис. 5), що підтверджує дієвість кожного методу зокрема та їх поєднання для підвищення достовірності стилевої атрибуції тексту.

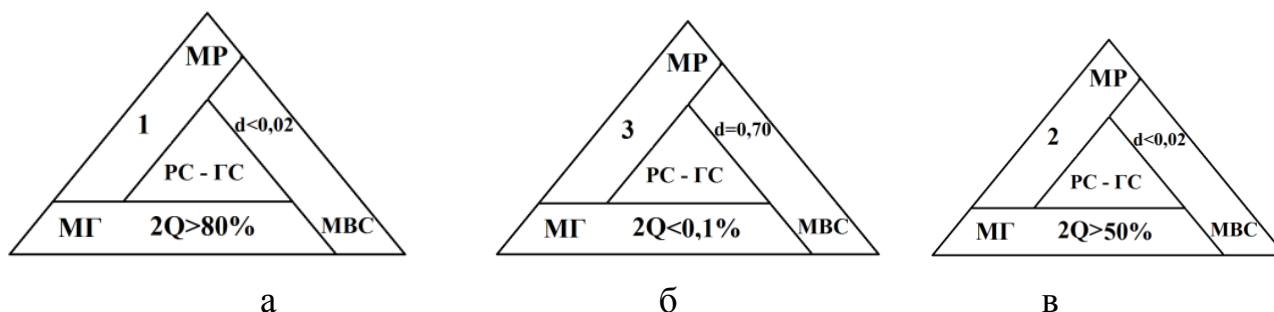


Рисунок 5 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі стилевої диференціації текстів за методом гіпотез, методом ранжування та методом визначення відстаней між стилями для зіставлених розмовного і газетного стилів за групою сонорних фонем для 3 випадків позиції фонем в слові: а) позиція фонем є невизначеною, б) фонема на початку слова, в) фонема в кінці слова. Істотні відмінності встановлено за МГ та МВС у випадку б.

Для підвищення достовірності та автоматизації стилевої, підстилевої та авторської атрибуції тексту, реалізовано вдосконалену статистичну модель визначення стилерозрізняльної здатності групи губних фонем за методом гіпотез при розмежуванні попарно зіставлених текстів поезії Дж. Г. Байрона (ПзБ), поезії Т. Мура (ПзМ), художньої прози Дж. Г. Байрона (Пр), драми Б. Шоу (Др), розмовного стилю (РС), газетного стилю (ГС) і наукового стилю (НС) (табл. 5).

Таблиця 5 – Результати визначення стилерозрізняльної здатності групи губних фонем

Стиль	ПзБ		ПзМ		Пр		Др		ГС		НС		РС	
	I	II	I	II	I	II	I	II	I	II	I	II	I	II
ПзБ			-	-	+	+	+	+	-	+	+	+	-	-
ПзМ	-	-			-	+	+	+	+	+	+	+	-	-
Пр	+	+	-	+			+	+	+	-	+	+	+	+
Др	+	+	+	+	+	+			+	-	-	+		
ГС	-	+	+	+	+	-	+	-			-	+		
НС	+	+	+	+	+	+	-	+	-	+			+	+
РС	-	-	-	-	+	+	+	+	-	+	+	+		

Встановлення істотних відмінностей (позначка + означає істотні відмінності; позначка - означає неістотні відмінності) за групою губних фонем, у наведених у табл. 5 зіставленнях для 2 випадків позиції фонем в слові (I – не враховуючи позицію фонем в слові; II – фонема на початку слова), показує рівень стилерозрізняльної здатності цієї групи фонем (істотні відмінності у 55 зіставленнях, неістотні – у 25 зіставленнях), що робить можливим розмежування текстів за цією групою фонем і тим самим забезпечує зменшення кількості груп фонем, підвищення рівня достовірності та автоматизації диференціації стилів.

Таким чином, розроблені методи і вдосконалені статистичні моделі дають змогу підвищити достовірність результатів, застосовуючи поєднання статистичних критеріїв, за якими отримано однаковий результат з довірчою ймовірністю 0,95 та рівень автоматизації стилевої, підстилевої та авторської атрибуції тексту.

ВИСНОВКИ

У дисертаційній роботі розв'язано актуальне наукове завдання підвищення достовірності диференціації стилів англійської мови на основі розроблених методів, моделей та програмних засобів.

1. Проведено аналіз використовуваних методів, моделей і засобів диференціації текстів, на основі чого сформульовано завдання та актуальність дисертаційного дослідження.

2. Розроблено метод комплексного аналізу диференціації фоностатистичних структур стилів англійської мови, складовими якого є статистичні методи ранжування, визначення відстаней між стилями та гіпотез з використанням потужного критерію Колмогорова-Смірнова. Застосоване поєднання статистичних методів дає змогу підтвердити результат, отриманий з довірчою ймовірністю 0,95,

кількома статистичними критеріями і тим самим підвищити достовірність диференціації стилів.

3. Розроблено багатофакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу, який ґрунтується на визначенні загальної стилевої маркованості зіставлених текстів. Загальна стилева маркованість досліджуваного тексту передбачає встановлення істотних відмінностей у трьох зіставленнях (стилів, підстилів, текстів різних авторів) з довірчою ймовірністю 0,95, що уможливило здійснення стилевої, підстилевої та авторської атрибуції досліджуваного тексту з більшою достовірністю.

4. Вдосконалено статистичну модель стилевої, підстилевої та авторської диференціації текстів за методом гіпотез і ранжування, яка визначає кількість груп фонем, за якими встановлено істотні відмінності з довірчою ймовірністю 0,95 між попарно зіставленими стилями. Розмежування зіставлених вибірок за декількома групами фонем забезпечує більшу достовірність визначення диференційних статистичних характеристик стилів.

5. Вдосконалено статистичну модель визначення стилерозрізняльної здатності груп приголосних фонем досліджуваних стилів, яка уможливило зменшення кількості груп приголосних фонем, за якими розмежовуються стилі, що забезпечує підвищення рівня автоматизації диференціації стилів.

6. Розроблено програмні засоби для автоматизації диференціації фоно-статистичних структур стилів англійської мови на мові програмування Java, яка забезпечує кросплатформність програмного засобу. Побудовано структуру системи диференціації стилів, яка ґрунтується на модульному принципі, що дає змогу швидко модифікувати і вдосконалювати програмну систему. Розроблено алгоритми для реалізації методів та моделей диференціації фоно-статистичних структур стилів англійської мови.

7. Результати виконаного дослідження використано як методичне забезпечення для дисциплін “Основи прикладної лінгвістики” та “Контрастивна стилістика англійської та української мов” у навчальному процесі кафедри прикладної лінгвістики Національного університету “Львівська політехніка”.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Khomytska I., Teslyuk V., Kryvinska N., Bazylevych I. Software-based approach towards automated authorship acknowledgement—chi-square test on one consonant group // *Electronics*. 2020. Vol. 7:1138. (Scopus)
2. Khomytska I., Teslyuk V., Bazylevych I., Shylinska I. Approach for minimization of phoneme groups in authorship attribution // *International Journal of Computing*. 2020. Vol. 19 (1). P. 55–62. (Scopus)
3. Khomytska I., Teslyuk V., Holovatyuk A., Morushko O. Development of methods, models and means for the author attribution of a text // *Eastern-European Journal of Enterprise Technologies*. Kharkiv, 2018. Is. 3/2 (93). P. 41–46. (Scopus)
4. Khomytska I., Teslyuk V. The method of statistical analysis of the scientific, colloquial, belles-lettres and newspaper styles on the phonological

- level // *Advances in Intelligent Systems and Computing* / Natalia Shakhovska editor. Lviv, 2016. Vol. 512. P. 149–163. (Scopus)
5. Khomytska I., Teslyuk V. Authorship and style attribution by statistical methods of style differentiation on the phonological level // *Advances in Intelligent Systems and Computing III* / Natalia Shakhovska editor. Lviv, 2018. Vol. 871. P. 105–118. (Scopus)
 6. Khomytska I., Teslyuk V. Statistical models for authorship attribution // *Advances in Intelligent Systems and Computing III* / Natalia Shakhovska editor. Lviv, 2020. Vol. 1080. P. 579–592. (Scopus)
 7. Хомицька І. Ю., Теслюк В. М. Статистичний аналіз англійських поетичних текстів // *Науковий вісник НЛТУ України*. Львів, 2015. Вип. 25.2. С. 350–356. (index Copernicus)
 8. Хомицька І. Ю., Теслюк В. М. Модель статистичного аналізу процесу функціонування груп англійських приголосних фонем у системі функціональних стилів // *Науковий вісник НЛТУ України*. Львів, 2015. Вип. 25.3. С. 364–369. (index Copernicus)
 9. Хомицька І. Ю., Теслюк В. М. Метод статистичного аналізу функціональних стилів англійської мови на фонологічному рівні // *Науковий вісник НЛТУ України*. Львів, 2015. Вип. 25.6. С. 322–328. (index Copernicus)
 10. Хомицька І. Ю., Теслюк В. М., Береговський В. В. Метод комплексного аналізу диференціації фоностатистичних структур стилів англійської мови // *Науковий вісник НЛТУ України*. Львів, 2019. Вип. 29.6. С. 140–143. (index Copernicus)
 11. Хомицька І. Ю., Теслюк В. М., Береговський В. В. Математичні метод і модель диференціації фоностатистичних структур авторського стилю // *Науковий вісник НЛТУ України*. Львів, 2019. Вип. 29.7. С. 155–159. (index Copernicus)
 12. Хомицька І. Ю., Теслюк В. М. Метод статистичного аналізу художнього стилю на фонологічному рівні // *Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі*. Львів, 2015. № 814. С. 326–334.
 13. Хомицька І. Ю., Теслюк В. М. Статистичний метод визначення стилерозмежувальної здатності групи губних фонем в системі англійських стилів // *Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі*. Львів, 2016. № 854. С. 240–248.
 14. Хомицька І. Ю., Теслюк В. М. Метод статистичного аналізу текстів драматургії Б. Шоу і розмовного стилю // *Вісник Національного університету «Львівська політехніка». Комп'ютерні науки та інформаційні технології*. Львів, 2015. № 826. С. 161–169.
 15. Khomytska I., Teslyuk V. Specifics of phonostatistical structure of the scientific style in English style system // *Computer Science and Information Technologies (CSIT'2016): Proceedings of the XIth Scientific and Technical Conference*. Lviv, 2016. P. 129–131. (Scopus)
 16. Khomytska I., Teslyuk V. Modelling of phonostatistical structures of English backlingual phoneme group in style system // *The Experience of Designing and*

- Application of Computer Aided Design Systems in Microelectronics (CADMS'2017)*: Proceedings of the 14th International Conference. Polyana, 2017. P. 324–327. (Scopus)
17. Khomytska I., Teslyuk V. Modelling of phonostatistical structures of the colloquial and newspaper styles in English sonorant phoneme group // *Computer Science and Information Technologies (CSIT'2017)*: Proceedings of the XIIth Scientific and Technical Conference. Lviv, 2017. P. 67–70. (Scopus)
 18. Khomytska I., Teslyuk V. Authorship attribution by differentiation of phonostatistical structures of styles // *Computer Science and Information Technologies (CSIT'2018)*: Proceedings of the XIIIth Scientific and Technical Conference. Lviv, 2018. P. 5–8. (Scopus)
 19. Khomytska I., Teslyuk V. The software for authorship and style attribution // *The Experience of Designing and Application of Computer Aided Design Systems (CADMS'2019)*: Proceedings of the 15th International Conference. Polyana, 2019. Vol. 2. P. 23–26. (Scopus)
 20. Khomytska I., Teslyuk V. Mathematical methods applied for authorship attribution on the phonological level // *Computer Science and Information Technologies (CSIT'2019)*: Proceedings of the XIVth Scientific and Technical Conference. Lviv, 2019. P. 7–11. (Scopus)
 21. Khomytska I., Teslyuk V., Kryvinska N., Beregovskiy V. The nonparametric method for differentiation of phonostatistical structures of authorial style // *Procedia Computer Science*: Proceedings of the 10th International Conference on Emerging Ubiquitous Systems and Pervasive Networks. Coimbra, Portugal, 2019. Vol. 160. P. 38–45. (Scopus)
 22. Khomytska I., Teslyuk V. The multifactor method applied for authorship attribution on the phonological level // *Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020)*. Vol. I: Main Conference, Lviv, Ukraine, April 23–24, 2020. CEUR Workshop Proceedings 2604, CEUR-WS.org 2020. P. 189–198. (Scopus)
 23. Хомицька І. Ю., Теслюк В. М., Лабінська Л. С. Програмна система авторської атрибуції текстів на фонологічному рівні // *Проблеми та перспективи розвитку економіки і підприємництва та комп'ютерних технологій в Україні* : зб. тез XIV наук.-практ. конф. Львів, 2018. С. 15–16.

АНОТАЦІЇ

Хомицька І. Ю. Методи та засоби диференціації фоностатистичних структур функціональних стилів англійської мови. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 10.02.21 – структурна, прикладна та математична лінгвістика. – Національний університет “Львівська політехніка” Міністерства освіти і науки України, Львів, 2021.

Дисертаційна робота присвячена розв’язанню актуального наукового завдання – підвищення достовірності диференціації фоностатистичних структур стилів англійської мови. У дисертації розроблено метод комплексного аналізу

диференціації фоностатистичних структур стилів англійської мови, який дає змогу підвищити достовірність диференціації фоностатистичних структур стилів англійської мови. Розроблено багатофакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу. Метод забезпечує підвищення достовірності здійснення авторської атрибуції тексту за його загальною стилевою маркованістю. Вдосконалено статистичну модель стилевої, підстилевої та авторської диференціації за методом гіпотез та методом ранжування, яка уможливорює підвищення достовірності визначення диференційних ознак стилів, підстилів та текстів різних авторів. Вдосконалено статистичну модель визначення стилерозрізняльної здатності груп приголосних фонем, яка дає змогу зменшити кількість груп приголосних фонем та підвищити рівень автоматизації диференціації текстів за групою приголосних фонем з найвищою стилерозрізняльною здатністю. Розроблено програмну систему диференціації стилів англійської мови на фонологічному рівні та наведено результати дослідження.

Ключові слова: диференціація фоностатистичних структур стилів, функціональний стиль, статистична модель, стилева атрибуція, авторська атрибуція, програмне забезпечення.

Хомицкая И. Ю. Методы и средства дифференциации фоностатистических структур функциональных стилей английского языка. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 10.02.21 – структурная, прикладная и математическая лингвистика. – Национальный университет “Львовська политехніка” Министерства образования и науки Украины, Львов, 2021.

Диссертационная работа посвящена решению актуального научного задания – повышения достоверности дифференциации фоностатистических структур стилей английского языка. В диссертации разработан метод комплексного анализа дифференциации фоностатистических структур стилей английского языка, который дает возможность повысить достоверность дифференциации фоностатистических структур стилей английского языка. Разработан многофакторный метод определения степени действия факторов стиля, подстиля и авторской манеры изложения. Метод обеспечивает повышение достоверности осуществления авторской атрибуции текста за его общей маркированностью. Усовершенствовано статистическую модель стилевой, подстилевой и авторской дифференциации методом гипотез и методом ранжирования, которая дает возможность повысить достоверность определения дифференциальных признаков стилей, подстилей и текстов разных авторов. Усовершенствовано статистическую модель определения стилеразличительной способности групп согласных фонем, которая дает возможность уменьшить количество групп согласных фонем и повысить уровень автоматизации дифференциации текстов по группам согласных фонем с наивысшей стилеразличительной способностью. Разработана программная система дифференциации стилей английского языка на фонологическом уровне и приведены результаты исследования.

Ключевые слова: дифференциация фоностатистических структур стилей, функциональный стиль, статистическая модель, стилевая атрибуция, авторская атрибуция, программное обеспечение.

Khomytska I. Yu. The methods and means of differentiation of phonostatistical structures of English functional styles. – On the rights of the manuscript.

Thesis for the degree of candidate of technical sciences, speciality 10.02.21 – Structural, Applied and Mathematical Linguistics. – Lviv Polytechnic National University, Ministry of Education and Science of Ukraine, Lviv, 2021.

The thesis is devoted to the topical scientific problem of development of the mathematical methods and models that can improve the test validity of differentiation of phonostatistical structures of English functional styles.

In the introduction, the relevance of the thesis theme is substantiated and the purpose is stated. The information about the program's testing, and author's contribution is given. In the first part of the thesis, the current state of the style differentiation problem has been analyzed. The results of the analysis have shown that the task of the test validity enhancement should be solved. Consequently, a method based on a combination of statistical methods and a multifactor method of determination of style factor and authorial factor effect should be developed to improve the test validity of style differentiation.

In the second part, the method of complex analysis of differentiation of phonostatistical structures of styles has been developed. It is based on the proposed combination of the hypothesis, the ranking and the style distance determination methods. The method makes it possible to improve the test validity of style differentiation. The developed multifactor method of determination of style and authorial factor effect allows us to improve the test validity of authorship attribution.

In the third part, the statistical model of style and authorial differentiation by the hypothesis and ranking methods and the statistical model of determination of style-differentiating capability of a consonant group have been improved. The models enhance the test validity of style and authorship attribution and level of automation of text differentiation.

In the fourth part, the software for style and authorial differentiation has been developed. The software is based on the developed methods using the Java programming language. According to the obtained results, the essential differences have been established between compared in pairs the belles-lettres, colloquial, newspaper, publicist and scientific functional styles. Authorship attribution has been done for the texts by G. G. Byron, T. Moore, E. Bronte, W. Thackeray, B. Obama, D. Trump, S. Logan and D. Webster.

Key words: differentiation of phonostatistical structures of styles, functional style, statistical model, style attribution, authorship attribution, software.

Підписано до друку 17.03.2021 р.

Формат 60x90/¹⁶. Папір офсетний.

Друк цифровий. Умовн. друк. арк. 1,4.

Гарнітура Times New Roman.

Наклад 100 прим. Зам. № 142072.

Надруковано в ТзОВ «ЗУКЦ»

79011, м. Львів, вул. Вітовського, 25/10

тел.: +38 (032) 297-06-76

Свідоцтво суб'єкта видавничої справи ДК №408 від 09.04.2001