

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА ПОЛІТЕХНІКА»

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Кваліфікаційна наукова
праця на правах рукопису

Хомицька Ірина Юріївна

УДК 519.76:811.111'32(043.5)

ДИСЕРТАЦІЯ

**МЕТОДИ ТА ЗАСОБИ ДИФЕРЕНЦІАЦІЇ ФОНОСТАТИСТИЧНИХ
СТРУКТУР ФУНКЦІОНАЛЬНИХ СТИЛІВ АНГЛІЙСЬКОЇ МОВИ**

10.02.21 – Структурна, прикладна та математична лінгвістика
«05 – Технічні науки»

Подається на здобуття наукового ступеня кандидата технічних наук

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Хомицька І. Ю.

Науковий керівник

Теслюк Василь Миколайович,
доктор технічних наук, професор

Ідентичність усіх примірників дисертації

ЗАСВІДЧУЮ:

Учений секретар спеціалізованої
вченої ради Д 35.052.05

/Р.А. Бунь/

Львів – 2020

АНОТАЦІЯ

Хомицька І. Ю. Методи та засоби диференціації фоностатистичних структур функціональних стилів англійської мови. – Кваліфікаційна наукова праця 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, 2020.

Спеціалізована вчена рада Д 35.052.05 Національного університету «Львівська політехніка» Міністерства освіти і науки України, Львів.

Дисертацію присвячено підвищенню достовірності диференціації фоностатистичних структур стилів англійської мови на основі розроблених і вдосконалених методів, моделей та програмних засобів.

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. Подано 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, Р. Колера, 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. Крім того, поєднання потужних параметричних і непараметричних критеріїв диференціації стилів забезпечує підвищення рівня достовірності результатів з довірчою ймовірністю 0,95. Оскільки ідея поєднання статистичних методів є відсутньою в

існуючих дослідженнях, актуальним є розроблення методу комплексного аналізу диференціації стилів.

Проведений аналіз використовуваних методів показав, що в існуючих дослідженнях відсутній багатофакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу. Ідея методу полягає у визначенні загальної стилевої маркованості досліджуваного тексту. Визначення загальної стилевої маркованості досліджуваного тексту певного автора, враховуючи значення істотних відмінностей зумовлених трьома факторами (стилю, підстилю, авторської манери викладу), робить можливим підвищення достовірності авторської диференціації текстів з довірчою ймовірністю 0,95 та дає суттєву інформацію про характерні особливості авторського стилю. Тому актуальною складовою для досягнення мети дисертаційного дослідження є розроблення багатофакторного методу.

На основі проведеного аналізу існуючих моделей стилевої диференціації стилів слід зазначити вагомий доробок у розробленні 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 стилевої диференціації, 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 розмежовувати стилі за цією групою фонем.

Аналіз засобів диференціації текстів показав, що ефективними є програмні системи від NeoNeuro. 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 програмою системи авторської, підстилевої, і стилевої атрибуції тексту на фонологічному рівні.

На основі проведеного аналізу існуючих методів, моделей та програмних засобів диференціації стилів визначено основні завдання дослідження для підвищення достовірності та рівня автоматизації диференціації фоностатистичних структур стилів: розробити нові методи, засоби та вдосконалити статистичні моделі диференціації стилів на фонологічному рівні.

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

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

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 диференціація текстів за значенням загальної стилевої маркованості зіставлюваних текстів. Розроблений метод реалізується за схемою стиль→підстиль→автор, яка з більшою достовірністю забезпечує можливість встановлення загальної рівнодіючої факторів стилю, підстилю та авторської манери викладу і здійснення стилевої,

підстилевої та авторської атрибуції досліджуваного тексту за різницею значень загальної стилевої маркованості зіставляваних текстів з довірчою ймовірністю 0,95.

У третьому розділі побудовано модель статистичної системи фоностатистичних структур досліджуваних стилів (художнього, розмовного, газетного, публіцистичного та наукового) англійської мови, яка є аналогом системи функціональних стилів англійської мови. Складовими цієї статистичної системи 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.

З метою розв'язання завдання підвищення достовірності визначення диференційних ознак стилів, підстилів, та текстів різних авторів вдосконалено статистичну модель стилевої, підстилевої, та авторської диференціації текстів за методом гіпотез і ранжування, яка на відміну від існуючих, враховує позицію фонем у слові, визначає відношення кількості груп приголосних фонем, за якими встановлено істотні відмінності до загальної кількості груп приголосних фонем. Ідея диференціації стилів за кількістю груп приголосних фонем пропонується вперше. Велика кількість груп приголосних фонем (6–8), за якими тексти відрізняються істотно з довірчою ймовірністю 0,95, дає змогу з більшою достовірністю робити висновки про їх відмінність.

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 розмежовуються стилі, спрощує процес авторської атрибуції тексту, підвищує його достовірність і забезпечує вищий рівень автоматизації шляхом здійснення авторської атрибуції тексту за групою приголосних фонем з найбільшою стилерозрізняльною здатністю.

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, яка ґрунтується на модульному принципі, що дає змогу швидко модифікувати розроблене програмне забезпечення, підвищити рівень автоматизації та ефективність стилевої, підстилевої та авторської атрибуції тексту. На основі розробленої структури системи побудовано алгоритм функціонування програми автоматизації диференціації функціональних стилів англійської мови 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. Побудовано структуру класів ПЗ системи диференціації стилів. Програмна реалізація розроблених методів і вдосконалених моделей забезпечила одержання наведених нижче результатів.

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 за розробленим методом комплексного аналізу диференціації фоностатистичних структур стилів англійської мови.

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

Список публікацій здобувача 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*. July 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., Holovatyy 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 НУ «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 НУ «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, 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., Beregovskyi 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.

SUMMARY

Khomytska I. Yu. The Methods and Means of Differentiation of Phonostatistical Structures of English Functional Styles. – Qualification scientific work on the rights of the manuscript.

Thesis for a candidate of technical sciences degree in specialty 10.02.21 – Structural, Applied and Mathematical Linguistics. – Lviv Polytechnic National University, Ministry of Education and Science of Ukraine, Lviv, 2020.

The specialized board D 35.052.05 of Lviv Polytechnic National University, Ministry of Education and Science of Ukraine, Lviv.

The thesis is devoted to improvement of test validity of differentiation of phonostatistical structures of English functional styles on the basis of the developed and improved mathematical methods, models and software.

In the introduction, relevance of the theme of the thesis has been substantiated, the purpose and main tasks of the research have been formulated, the scientific novelty and practical application of the work have been stated, the connection of the research with research themes has been shown. Information about program's testing, author's contribution and publications has been given.

In the first part, the existing methods, models and software of style differentiation as well as authorship attribution of a text have been analysed.

Application of the probabilistic methods and the mathematical, statistical methods is advisable in research of a language which is a probabilistic system (works by G. Altmann, R. Koler, V. Perebyjnis, V. Levytskiy). The analysis of the research on the phonological level shows that the level of test validity is not sufficient because the texts are not differentiated by a powerful nonparametric Kolmogorov-Smirnov's test.

In addition to this, a combination of powerful parametric and nonparametric tests of style differentiation secures enhancement of a test validity level up to 95%. As the researchers have not used a combination of statistical methods, it is topical to develop the method of complex analysis of style differentiation.

The conducted analysis of the existing methods has shown that the multifactor method of determination of style and authorial factor effect has not been applied. The method is based on determination of general style markedness of a researched text. As the method gives essential information about individual peculiarities of an authorial style and makes it possible to enhance a test validity level of style differentiation up to a 95%, it is topical to develop the multifactor method. The results of the existing style differentiation models allow us to draw a conclusion that a valuable contribution in the development of statistical models of style differentiation has been made by V. Perebyinis with the help of the Student's t-test. However, for enhancement of test validity of style differentiation it is necessary to improve the existing statistical models using a powerful nonparametric Kolmogorov-Smirnov's test and to differentiate styles by the number of phoneme groups in which the essential differences are established by the hypothesis and ranking methods. To enhance the level of test validity and automation of style and authorship attribution of a text, it is necessary to improve the statistical model of determination of style-differentiating capability of a consonant phoneme group of the researched styles on the basis of determination of the group with the highest style-differentiating capability and differentiate styles by this group.

Analysis of the existing software of style differentiation has shown that the programs from NeoNeuro are the most efficient. However, the programs process texts on the lexical-semantic level, the structure of which is less strict and it is more difficult to formalize it. Consequently, the test validity on the lexical-semantic level is lower than on the phonological one. Taking this into account, it is topical to develop the software of style and authorship attribution on the phonological level. On the basis of the conducted analysis of the existing methods, models and software the main tasks of the research have been determined to enhance reliability, test validity and level of automation of differentiation of phonostatistical structures of styles. The tasks are as follows: to develop new methods, software and improve the statistical models of style differentiation on the phonological level.

In the second part, the methods of differentiation of English styles have been developed. These are: the method of complex analysis of differentiation of phonostatistical

structures of styles and the multifactor method of determination of style and authorial factor effect. To improve reliability of differentiation of styles and texts by different authors, the method of complex analysis of differentiation of phonostatistical structures of styles has been developed. The method is based on a combination of the following statistical methods: the hypothesis method (the Student's t-test, the Kolmogorov-Smirnov's test, the chi-square test), the ranking method and the style distance determination method. Improvement of a test validity of style differentiation is ensured by the use of five methods, one of which is nonparametric with strength close to 1 (the Kolmogorov-Smirnov's test). The test validity level is 95%.

To improve test validity of style and authorship attribution of a text, a multifactor method of determination of style, genre and authorial factors effect has been developed. The method involves determination of general style markedness of a researched text (by the proposed formula). The peculiarity of the method consists of style differentiating by the values of general style markedness. The developed method is realized in the scheme: style → genre → author which makes it possible to improve the level of test validity, determine the resulting value of style, genre and authorial factors effect with a 95% confidence level and perform the authorship attribution of a researched text by difference of values of general style markedness of the compared texts.

In the third part, a system of phonostatistical structures of English styles (belles-lettres, colloquial, newspaper, publicist and scientific) has been built. This system is an analogue of the functional English style system. The constituents of this system are the improved models: 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 phoneme group.

To improve the test validity level for determining the style and authorial distinctive features, the statistical model of style and authorial differentiation by the hypothesis and ranking methods has been developed. The model takes into account a phoneme's position in a word and determines the ratio of the number of consonant phoneme groups by which the essential differences were established to the total number of consonant phoneme groups. It is the first time the styles are differentiated by a number of consonant phoneme

groups. A great number of consonant phoneme groups (6 – 8) in which the texts differ essentially with a 95% test validity level makes it possible to state the essential differences of styles with greater test validity.

To solve the problem of improvement of the level of automation of style differentiation, the statistical model of determination of style-differentiating capability of a consonant phoneme group has been improved. The model is based on the ratio of the number of comparisons in which the essential differences were established to the number of all comparisons. Reduction of the number of consonant phoneme groups, by which the texts are differentiated, simplifies the authorship attribution process, enhances its reliability and secures a higher level of automation in the group of consonant phonemes with the highest style-differentiating capability.

In the fourth part, the developed software for automated differentiation of phonostatistical structures of English functional styles has been described and the research results have been presented. The structure of the software of automated differentiation of phonostatistical structures of English functional styles has been developed. The structure is based on a modular principle which makes it possible to quickly modify the developed software, improve the automation level and efficiency of style and authorial attribution of a text. On the basis of the developed software structure, an algorithm of functioning of the program of automation of English functional style differentiation has been built on the phonological level whose elements are more strictly arranged and therefore the obtained on this level results have greater test validity. The developed software of the program is cross-platform because the program is developed on the object-oriented Java programming language. The structure of classes of the software of automated differentiation of phonostatistical structures of English functional styles has been built. The given below results have been obtained by the realized in the program developed methods and improved models.

Different degrees of difference have been determined in different consonant phoneme groups in a comparison of the belle-lettres, colloquial, newspaper, publicist and scientific styles. Great essential differences have been obtained in a comparison of the scientific (texts from technical physics (*Acta Physica Polonica*)) and colloquial (every-day lexis

(Free French Travel Phrasebook)) styles in the forelingual phoneme group by the developed method of complex analysis of differentiation of phonostatistical structures of styles.

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

**Scientific papers, in which the main scientific results of the dissertation
have been published**

1. Khomytska I., Teslyuk V., Kryvinska N., Bazylevych I. Software-based approach towards automated authorship acknowledgement—chi-square test on one consonant group // *Electronics*. July 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., Holovatyy 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. Khomytska I. Yu., Teslyuk V. M. Statystychnyy analiz anglijs'kyh poetychnykh tekstiv // *Naukovyy visnyk NLTU Ukrainy*. Lviv, 2015. Vyp. 25.2. S. 350–356. (index Copernicus)
8. Khomytska I. Yu., Teslyuk V. M. Model' statystychnoho analizu protsesu funktsionuvannia grup anglijs'kyh prygholosnykh fonem u systemi funktsional'nykh styliv // *Naukovyy visnyk NLTU Ukrainy*. Lviv, 2015. Vyp. 25.3. S. 364–369. (index Copernicus)
9. Khomytska I. Yu., Teslyuk V. M. Metod statystychnoho analizu funktsional'nykh styliv anglijs'koi movy na fonologichnomu rivni // *Naukovyy visnyk NLTU Ukrainy*. Lviv, 2015. Vyp. 25.6. S. 322–328. (index Copernicus)
10. Khomytska I. Yu., Teslyuk V. M., Beregovskyy V. V. Metod kompleksnoho analizu dyferentsiatsii fonostatystychnykh struktur styliv anglijs'koi movy // *Naukovyy visnyk NLTU Ukrainy*. Lviv, 2019. Vyp. 29.6. S. 140–143. (index Copernicus)
11. Khomytska I. Yu., Teslyuk V. M., Beregovskyy V. V. Matematychni metod i model' dyferentsiatsii fonostatystychnykh struktur avtors'koho stylu // *Naukovyy visnyk NLTU Ukrainy*. Lviv, 2019. Vyp. 29.7. S. 155–159. (index Copernicus)
12. Khomytska I. Yu., Teslyuk V. M. Metod statystychnoho analizu khudozhn'oho stylu na fonologichnomu rivni // *Visnyk Natsional'noho universytetu "L'vivs'ka politekhnika"*. *Informatsijni systemy ta merezhi*. L'viv, 2015. № 814. S. 326–334.
13. Khomytska I. Yu., Teslyuk V. M. Statystychnyy metod vyznachennia stylerozmezhuval'noi zdatnosti grupy gubnykh fonem v systemi anglijs'kykh styliv // *Visnyk Natsional'noho universytetu "L'vivs'ka politekhnika"*. *Informatsijni systemy ta merezhi*. L'viv, 2016. № 854. S. 240–248.
14. Khomytska I. Yu., Teslyuk V. M. Metod statystychnoho analizu tekstiv dramaturgii B. Show i rozmovnoho stylu // *Visnyk Natsional'noho universytetu "L'vivs'ka politekhnika"*. *Kompyuterni nauky ta informatsijni tekhnologii*. L'viv, 2015. № 826. S. 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, 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., Beregovskyi 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. Khomytska I. Yu., Teslyuk V. M., Labins'ka L. S. Programna systema avtors'koi atrybutsii tekstiv na fonologichnomu rivni // *Problemy ta perspektyvy rozvytku ekonomiky i pidpryjemnytstva ta komp'uternyh tehnologij v Ukraini* : zb. tez XIV nauk.-prak. konf. L'viv, 2018. S. 15–16.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	21
ВСТУП	22
РОЗДІЛ 1. АНАЛІЗ МАТЕМАТИЧНИХ МЕТОДІВ, МОДЕЛЕЙ ТА ПРОГРАМНИХ ЗАСОБІВ ДЛЯ ДИФЕРЕНЦІАЦІЇ ФУНКЦІОНАЛЬНИХ СТИЛІВ АНГЛІЙСЬКОЇ МОВИ НА ФОНОЛОГІЧНОМУ РІВНІ	
1.1 Причини виникнення та історичний хід розвитку квантитативної фонології	29
1.2 Розвиток структуралізму – передумова виникнення математичної (квантитативної) лінгвістики	30
1.3 Аналіз статистичних моделей диференціації текстів для вивчення системно-структурної організації мови	32
1.4 Диференціація текстів у математичній лінгвістиці	35
1.5 Диференціація текстів у статистичній стилістиці та стилеметрії	38
1.6 Аналіз програмних засобів диференціації стилів	42
1.7 Висновки до розділу 1	44
РОЗДІЛ 2. РОЗРОБЛЕННЯ СТАТИСТИЧНИХ МЕТОДІВ ДЛЯ ДИФЕРЕНЦІАЦІЇ ФНОСТАТИСТИЧНИХ СТРУКТУР СТИЛІВ АНГЛІЙСЬКОЇ МОВИ	
2.1 Розроблення методу комплексного аналізу диференціації фоностатистичних структур стилів англійської мови	45
2.2 Використання методу гіпотез для диференціації фоностатистичних структур стилів англійської мови	47
2.3 Розроблення методу ранжування для диференціації фоностатистичних структур стилів англійської мови	58
2.4 Використання методу визначення відстаней між стилями для диференціації фоностатистичних структур стилів англійської мови	63
2.5 Розроблення багатofакторного методу для диференціації фоностатистичних структур стилів англійської мови	65
2.6 Висновки до розділу 2	68

РОЗДІЛ 3. МОДЕЛІ ТА АЛГОРИТМИ МОДЕЛІ СИСТЕМИ ФОНОСТАТИСТИЧНИХ СТРУКТУР ДОСЛІДЖУВАНИХ СТИЛІВ АНГЛІЙСЬКОЇ МОВИ

- | | |
|--|----|
| 3.1 Розроблення статистичних моделей – складових моделі системи
фоностатистичних структур стилів | 69 |
| 3.2 Розроблення алгоритму врахування позиції фонем в слові для моделі
системи фоностатистичних структур стилів | 76 |
| 3.3 Розроблення алгоритму визначення частотних характеристик груп
приголосних для моделі системи фоностатистичних структур стилів | 77 |
| 3.4 Побудова алгоритму опису рівнів мови та мовлення для моделі
системи фоностатистичних структур стилів | 81 |
| 3.5 Висновки до розділу 3 | 85 |

РОЗДІЛ 4. РОЗРОБЛЕННЯ ПРОГРАМНОЇ СИСТЕМИ ДИФЕРЕНЦІАЦІЇ ФОНОСТАТИСТИЧНИХ СТРУКТУР ФУНКЦІОНАЛЬНИХ СТИЛІВ АНГЛІЙСЬКОЇ МОВИ ТА РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

- | | |
|---|-----|
| 4.1 Розроблення структури системи диференціації фоностатистичних
структур функціональних стилів англійської мови | 87 |
| 4.2 Розроблення алгоритму функціонування системи диференціації
фоностатистичних структур функціональних стилів англійської мови | 89 |
| 4.3 Розроблення програмного забезпечення системи диференціації
фоностатистичних структур функціональних стилів англійської мови | 91 |
| 4.4 Розроблення інформаційного забезпечення системи диференціації
фоностатистичних структур функціональних стилів англійської мови | 96 |
| 4.5 Результати дослідження фоностатистичної структури розмовного
стилю | 97 |
| 4.6 Результати дослідження фоностатистичної структури художнього
стилю | |
| 4.6.1 Результати дослідження фоностатистичної структури поезії | 103 |
| 4.6.2 Результати дослідження фоностатистичної структури художньої
прози | 112 |

4.6.3 Результати дослідження фоностатистичної структури драми	116
4.7 Результати дослідження фоностатистичної структури газетного стилю	119
4.8 Результати дослідження фоностатистичної структури наукового стилю	125
4.9 Результати дослідження фоностатистичної структури авторського стилю	133
4.10 Висновки до розділу 4	147
ВИСНОВКИ	151
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	153
ДОДАТКИ	170
Додатки А.1 – А.64. Результати порівняння текстів за методом гіпотез	171
Додатки Б.1 – Б.8. Результати отримані за критерієм Пірсона для 8 груп приголосних фонем	235
Додатки В.1 – В.8. Графіки нормального розподілу імовірностей частот 8 груп приголосних фонем	243
Додаток Д. Графік залежності величини Δ від кількості порцій вибірки n	247
Додатки Е.1 – Е.6. Результати порівняння текстів за методами ранжування та визначення відстаней між стилями	248
Додаток Ж.1. Процес формування вибірки	254
Додаток Ж.2. Цикл формування порцій	254
Додаток Ж.3. Процес обчислення для кожної групи фонем	254
Додаток Ж.4. Процес обчислення за критерієм Пірсона	255
Додаток Ж.5. Процес обчислення за критерієм Стьюдента	255
Додаток Ж.6. Процес обчислення за критерієм Колмогорова-Смірнова	256
Додаток Ж.7. Процес обчислення за критерієм χ^2 -квадрат	256
Додаток К. Акт впровадження результатів дисертаційного дослідження	257
Додаток Л. Список публікацій здобувача за темою дисертації	258
Додаток М. Апробація роботи	262

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

Др – драма

ГС – газетний стиль

НС – науковий стиль

ПзБ – поезія Байрона

ПзМ – поезія Мура

Пр – художня проза

РС – розмовний стиль

ХС – художній стиль

МВС – метод визначення відстаней між стилями

МГ – метод гіпотез

МР – метод ранжування

СТ – критерій Стьюдента

КС – критерій Колмогорова-Смірнова

ХК – критерій хі-квадрат

ЗН – група задньоязикових фонем

ЩЛ – група щілинних фонем

ВСТУП

Актуальність теми. Широкомасштабне впровадження інформаційних технологій в усі сфери людської діяльності забезпечує глобалізацію світової спільноти. Глобалізація передбачає опрацювання великого обсягу інформації у текстовому поданні, 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 у статтях, що мають більш ніж одного автора); в офіційно-діловій сфері (документи); у сфері літератури (художні твори) та інше.

Встановлення авторства тексту ґрунтується на визначенні частоти вживання мовних одиниць, яка є диференційною характеристикою авторського стилю. Методи авторської атрибуції тексту розроблялися відповідно до рівня науково-технічного прогресу – від простих арифметичних дій до складніших кількісних методів, а від них – до ймовірнісних методів та методів математичної статистики. 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 ефективні статистичні методи для фонологічного рівня мови; Г. Альтманом, Р. Колером – визначено ефективні статистичні методи для кожного рівня мови зокрема; Д. Сегалом – сформульовано основи фонологічної статистики; В. Литвином, В. Висоцькою – розроблено методикау диференціації текстів за ключовими словами; О. Бісікало – розроблено методикау диференціації текстів на синтаксичному рівні; Ш. Аргамомом, М. Копелем – розроблено методикау встановлення авторства анонімних листів; Дж. Шлером – розроблено методикау автоматизованої диференціації текстів на лексико-семантичному рівні мови.

Зазначивши вагомий доробок попередніх досліджень, слід наголосити, що застосовані дослідниками методи дали змогу отримати різні значення достовірності авторської атрибуції тексту. Відповідно до існуючих результатів, найчастіше

дослідники вибирають рівень значущості 0,05, який є оптимальним і тексти різних авторів відрізняються з довірчою ймовірністю 0,95. Через те, що низка офіційних документів потребує вищого рівня надійності, його можна отримати, застосувавши поєднання ефективних статистичних методів, які дають однаковий результат на тому самому матеріалі. Крім цього, сучасні інформаційні технології дають змогу опрацьовувати великий обсяг інформації, що робить можливим підвищення параметрів достовірності. Тому, з огляду на сучасний стан досліджень, актуальним є наукове завдання підвищення достовірності диференціації текстів. Оскільки це завдання є метою дисертаційного дослідження, тема дисертації “Методи та засоби диференціації фоностатистичних структур функціональних стилів англійської мови” є актуальною.

Дисертаційне дослідження проведено на матеріалі текстів англійською мовою, бо це мова більшості видань, що входять до наукометричних баз Scopus та Web of Science, це мова міжнародних конференцій, на цій мові є найбільше літератури з інформаційних технологій, які функціонують у всіх суспільних сферах. Досягнення світової науки мають англомовні переклади. Доречно зазначити, що найбільше досліджень авторської атрибуції є на англійській мові, що дає змогу, використовуючи попередні здобутки, розширити спектр напрямів дослідження. У дослідження отримано результати за визначеним розподілом частот груп приголосних фонем. Приголосні фонемі кількісно переважають голосні фонемі, і тому вибір приголосних забезпечує достовірність результатів. Це підтверджено дослідженням Б. Бурдона, який встановив, що дентальні приголосні переважають в усіх мовах. Роботи П. Хула, Л. Бомб’єна, М. Пуплера, К. Мушамера та Б. Кунерта розкривають специфіку функціонування груп приголосних.

Результати дослідження одержано на матеріалі функціональних стилів англійської мови. Дослідження проведено у двох основних напрямках з метою визначення ефективності розмежування текстів методами математичної статистики при врахуванні дії часового чинника: 1) здійснення стилевої атрибуції текстів одного часового періоду; 2) різних часових періодів. У першому напрямі зіставлено тексти розмовного, газетного та наукового стилів. У другому напрямі порівняно

тексти підстилів художнього стилю: поезії, художньої прози і драми. Художній стиль зіставляється з розмовним, газетним і науковим. У плані авторської атрибуції порівняно тексти художнього та публіцистичного стилів, які належать до одного часового періоду.

Зв'язок роботи з науковими програмами, планами, темами. Тема дисертації відповідає науковому напрямку 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, підстилів 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 уможлиблює зменшення кількості груп приголосних фонем, та забезпечує 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 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, підстилів і стилів та робить можливим 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, робить можливим визначення групи приголосних фонем з найвищою стилерозрізняльною здатністю і, тим самим, зменшує кількість груп приголосних фонем, за якими диференціюються 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, які уможливають ефективніше 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” та “Контрастивна стилістика англійської та української мов” у навчальному процесі кафедри прикладної лінгвістики Національного університету “Львівська політехніка”, що підтверджено відповідним актом (Додаток К).

Особистий внесок здобувача. Всі наукові результати теоретичних і практичних досліджень, які представлено 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 [183]; статистична модель стилевої диференціації наукового, розмовного, газетного і художнього стилів за методом гіпотез та результати досліджень [121, 122, 153]; багатofакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу та алгоритм його реалізації [148, 171]; статистична модель визначення стилерозрізняльної, підстилерозрізняльної та авторорозрізняльної здатності груп приголосних фонем, результати досліджень [181, 169, 126, 123, 127, 146]; статистична модель авторської диференціації за методом гіпотез і ранжування, результати досліджень [183, 182]; статистична модель підстилевої диференціації за методом гіпотез, результати досліджень [153, 143]; статистична модель стилевої диференціації за групою середньоязикових фонем за методом гіпотез, результати досліджень [165]; статистична модель стилевої диференціації за методами гіпотез, ранжування і визначення відстаней між стилями, результати досліджень [160]; алгоритм, структура та програмне забезпечення системи диференціації фоностатистичних структур стилів, результати досліджень [180]; алгоритм визначення стилерозрізняльної здатності груп приголосних фонем, результати досліджень [170, 184, 168]; алгоритм авторської атрибуції тексту, результати досліджень [112].

Апробація роботи. Основні теоретичні положення та практичні результати дисертаційної роботи доповідалися і обговорювалися на: 11-й Міжнародній науково-технічній конференції “Комп’ютерні науки та інформаційні технології”, Національного університету “Львівська політехніка”, 2016; 14-й Міжнародній науково-технічній конференції “Досвід розробки та застосування систем автоматизованого проектування в мікроелектроніці”, Національного університету “Львівська політехніка”, 2017; 12-й Міжнародній науково-технічній конференції “Комп’ютерні науки та інформаційні технології”, Національного університету “Львівська політехніка”, 2017; 13-й Міжнародній науково-технічній конференції “Комп’ютерні науки та інформаційні технології”, Національного університету “Львівська політехніка”, 2018; 15-й Міжнародній науково-технічній конференції “Досвід розробки та застосування систем автоматизованого проектування”, Національного університету “Львівська політехніка”, 2019; 14-й Міжнародній

науково-технічній конференції “Комп’ютерні науки та інформаційні технології”, Національного університету “Львівська політехніка”, 2019; 4-й Міжнародній науково-технічній конференції “Комп’ютерна лінгвістика та інтелектуальні системи”, Національного університету “Львівська політехніка”, 2020; 14-й науково-практичній конференції “Проблеми та перспективи розвитку економіки і підприємництва та комп’ютерних технологій в Україні”, Національного університету “Львівська політехніка”, 2018.

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 міжнародної наукометричної бази Scopus та Web of Science.

Структура та обсяг роботи. Дисертація складається з анотації, вступу, 4-х 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.

РОЗДІЛ 1. АНАЛІЗ МАТЕМАТИЧНИХ МЕТОДІВ, МОДЕЛЕЙ ТА ЗАСОБІВ ДЛЯ ДИФЕРЕНЦІАЦІЇ ФУНКЦІОНАЛЬНИХ СТИЛІВ АНГЛІЙСЬКОЇ МОВИ НА ФОНОЛОГІЧНОМУ РІВНІ

1.1 Причини виникнення та історичний хід розвитку квантитативної фонології

Основою всієї наукової роботи є твердження, що світ є впорядкованою і доступною для пізнання сутністю [1]. Це твердження А. Ейнштейна, природно, правомірне і при дослідженні мови. Адже мова теж є впорядкованою і доступною для пізнання сутністю. Мова є системно-структурним утворенням. Системно-структурний характер мови найбільш чітко виражений на фонологічному рівні. Наявність постійного, а не змінного складу мовних одиниць, наявність зв'язків між мовними одиницями проти безладної та хаотичної їх взаємодії – це саме ті фактори, які не лише дають змогу застосовувати методи математичної статистики, але й зумовлюють необхідність їх застосування у лінгвістичних дослідженнях. Встановлені статистичні характеристики вносять ясність, чіткість та однозначність у трактування тих чи інших мовних явищ і дають змогу зробити важливі теоретичні висновки щодо закономірностей функціонування мовної системи загалом.

Проблема виявлення частотності характерної для певної знакової одиниці з давніх давен привертала увагу дослідників текстів. Першу спробу дати точний опис мовним явищам зроблено у роботах відомого стародавнього індійського вченого Паніні (5 ст. до н.е.) [2]. На ранніх етапах дослідження не розрізнялися не тільки фонема і звук, але й фонема і буква. Дослідників цікавила практична сторона стабільності частоти вживання певної знакової одиниці. Закономірність повторюваності певних знакових одиниць використовувалась у шифрах і дешифруванні впродовж багатьох століть, але лише у дев'ятнадцятому столітті отримала своє теоретичне обґрунтування. У кінці XIX ст., відповідно до закону сформульованого Б. Бурдоном [3], встановлено переважання дентальних звуків у всіх мовах. Важливість урахування статистичних характеристик при вивченні

закономірностей мовних одиниць задекларовано у першому частотному словнику укладеному Дж. В. Кедінгом [4]. У середині ХХ ст. Ванг і Крофорд вивчають специфіку функціонування англійських приголосних фонем. Роботи виконано на високому лінгвістичному рівні. Вони містять вагому лінгвістичну інтерпретацію статистичних даних і до сьогодні не втрачають актуальності [5].

1.2 Розвиток структуралізму – передумова виникнення математичної (квантитативної) лінгвістики

У кінці ХІХ – на початку ХХ ст. розробляються основні поняття і засади структурної лінгвістики, яка зумовила появу таких напрямів як математична (квантитативна) лінгвістика, лінгвістична статистика та комп'ютерна лінгвістика.

Основними передумовами виникнення досліджень напряму структуралізму було виявлення системно-структурної організації мови. Для точного, формального опису мовних об'єктів, в якому застосовується математичний апарат, досліджувані мовні об'єкти повинні бути системно організованими.

В історії мовознавства системний характер мови вперше визнано видатним німецьким дослідником Гумбольдтом [6]. Вирішальну роль у застосуванні системно-структурного підходу до мови відіграв видатний представник Женевської лінгвістичної школи Фердинанд де Соссюр, який вважав мову знаковою системою, формою, тобто структурою, організацією звукової матерії [7].

Ідеї Фердинанда де Соссюра розвинули інші представники структуралізму. Зокрема, представник датського структуралізму Луї Єльмслев [8] та представник казанської лінгвістичної школи Бодуен де Куртене, який називав мову узагальненою конструкцією. Бодуен де Куртене вважав, що необхідно частіше застосувати у мовознавстві кількісне, математичне мислення і наближувати його до точних наук. Використовуючи складні формули, він успішно застосовує кількісні методи у дослідженнях фонетики [9].

Виявлення системно-структурних зв'язків між мовними одиницями, зокрема фонемами, розподіл частот яких досліджується у даній роботі, передбачає трактування фонему як звуковий тип. Такий підхід до розуміння фонему лежить в основі теорії фонему розробленої Л. В. Щербою – засновником Ленінградської фонологічної школи (1912). Згідно з цією теорією фонема є односторонньою одиницею, що полегшує вивчення системно-структурних зв'язків у фонологічній системі [10]. Саме з цієї причини фонологічний рівень раніше за інші рівні досліджується статистичними методами.

В українському мовознавстві Є. К. Тимченко визначає фонему як звуковий тип, що об'єднує групу історично близьких діалектних варіантів у роботі «Курс історії українського язика. Вступ і фонетика» (1927) [11].

У дисертаційному дослідженні використовується трактування фонему розроблене Ленінградською фонологічною школою. Фонема – звукотип, мінімальна одиниця звукової будови мови, яка слугує для розпізнання та розрізнення значенневих одиниць – морфем і слів [12].

Російський дослідник знакової теорії мови В. М. Солнцев вважає мову системою, якій властива як форма, так і субстанція і трактує фонему як матеріальну основу всіх двосторонніх одиниць мови і мовлення. Фонема є пучком диференційних ознак, за якими об'єднується в клас низка фізично подібних і функційно тотожних звуків [13]. Пучок диференційних ознак є незмінним, інваріантним в усіх її реалізаціях на рівні мовлення, тобто у кожному варіанті – формі буття фонему. Таким чином, визначення частотності конкретних реалізацій певної фонему в мовленні, дає змогу отримати дані про специфіку функціонування самої фонему. Таке протиставлення фонему – абстрактної сутності, інваріанта своєму варіантові є протиставленням двох рівнів – мови і мовлення, парадигматики і синтагматики.

Про роль системності у мові слушно завважує видатний теоретик українського мовознавства О. С. Мельничук, який пов'язує прогрес у мові із посиленням системності, а регрес – із послабленням системності [14].

Складність фонологічної системи проаналізовано у роботах: К. Купа, Е. Марсіко, Ф. Пелегріно [15], П. Хула, Л. Бомб'єна, М. Пуплера, К. Мушамера, Б. Кунерта [16]. У цих роботах зазначено, що визначення диференційних ознак, за якими фонема протиставляється іншим фонемам не має однозначного трактування, що спричинює відсутність уніфікованої класифікації.

Основним нерозв'язаним завданням досліджень структуралізму є те, що методика 'система в системі' не застосовувалась у достатній мірі. Поділ системи на підсистеми є науково обгрунтованим у випадку вивчення мови, яка має не жорстко детерміновану, а ймовірнісну природу і потребує докладнішого вивчення підсистем.

1.3 Аналіз статистичних моделей диференціації текстів для вивчення системно-структурної організації мови

Дослідження ймовірнісного характеру мови є складною, багатоаспектною проблемою, яка не втрачає своєї актуальності. Різноманітність лінгвістичних та екстралінгвістичних чинників, що зумовлюють вживання мовної одиниці в мовленнєвому ланцюгу, часто важко ідентифікувати на рівні інтуїтивних узагальнень та елементарних аналогій. Тому необхідно застосовувати метод строгої аналогії, який лежить в основі моделювання. Важливим етапом у розвитку структурної лінгвістики є виникнення поняття лінгвістичної моделі. Модель є потужним апаратом в руках дослідника. Конструювання структури певної мовної системи не лише відображає загальні закономірності функціонування цієї системи, але і є об'єктивним критерієм істинності наших знань про мову. Моделі побудовані з використанням результатів одержаних методами математичної статистики дають змогу оцінити мовні процеси з деякою довірчою ймовірністю.

Аналітичні та синтетичні властивості людського мозку дають змогу систематизувати, класифікувати та моделювати мовні об'єкти і явища. Мовні об'єкти і явища конструюються за певною схемою – лінгвістичною моделлю, яка репрезентує і описує об'єкт дослідження під певним кутом зору. Кожна модель є продуктом абстракції та інваріантом стосовно її конкретних варіантних видозмін.

Модель, яка репрезентує мовну систему, є моделлю-системою, що складається з моделей-підсистем (моделі-підсистеми фонологічного рівня, моделі-підсистеми лексичного рівня та моделей інших рівнів мови). З усіх рівнів мови фонологічний є найбільш впорядкованим і систематизованим. Його найлегше формалізувати. Тому більшість перших моделей репрезентували специфіку функціонування одиниць фонологічного рівня.

Серед вагомих робіт присвячених моделюванню варто зазначити наукові роботи М. Д. Андрєєва, який досліджував структурно-імовірнісні опозиції, імовірнісну диференційну ознаку в плані структурно-імовірнісного моделювання мови [17]. Г. В. Єрмоленко досліджує специфіку імовірнісно-статистичної моделі [17]. У зазначених наукових роботах моделювання розглядається на рівні певного компонента системи виокремленого з певною метою.

Фонологічна система та структура є об'єктами моделювання у праці І. Ревзіна. Дослідник на основі класифікації звуків у певний історичний період побудував парадигматичну звукову модель. У дослідженні І. І. Ревзіна звуки розглядаються під структурним кутом зору. Звуки є елементами фонологічної системи, в якій вони займають визначене за їхніми диференційними ознаками місце [18]. Р. Г. Піотровський моделює фонологічні системи та визначає методи їх співставлення. Дослідник слушно зауважив, що фонологічна модель розкриває ті фонематичні явища, які без застосування точних методів є непомітними для лінгвістичної інтуїції фонематичними сутностями [19]. Всебічне висвітлення результатів застосування статистичних методів на парадигматичному рівні мови представлено у монографії "Основи фонологічної статистики" Д. М. Сегала. Побудована модель статистичної структури фонологічного рівня є передумовою рівноваги фонологічних систем. Рівновага фонологічних систем дає змогу отримати з мінімальної кількості фонологічних ознак максимальну кількість опозицій. [20].

Статистичні моделі фонологічного рівня широко представлено у дослідженнях В. І. Перебийніс. Дослідниця вважає, що статистичні моделі відбивають взаємовідношення кількісних та якісних характеристик фонологічного рівня. На її думку, статистичною моделлю можна вважати частотні характеристики

фонем чи груп фонем виокремлених за певними диференційними ознаками [21]. Зазначивши вагомий доробок дослідниці у розроблення статистичних моделей стилевої диференціації, які ґрунтуються на застосуванні параметричного критерію (критерію Стьюдента), слід зауважити, що для підвищення достовірності стилевої диференціації необхідно вдосконалити існуючі статистичні моделі, використовуючи потужний непараметричний критерій Колмогорова-Смірнова та диференціюючи стилі за кількістю груп приголосних фонем, за якими встановлено істотні відмінності за методом гіпотез та ранжування. Вдосконалення статистичної моделі визначення стилерозрізняльної здатності груп приголосних фонем досліджуваних стилів здійснено на основі визначення групи фонем з найбільшою стилерозрізняльною здатністю та розмежуванням стилів за цією групою фонем. Зменшення груп приголосних фонем забезпечує підвищення рівня автоматизації стилевої, підстилевої та авторської атрибуції тексту.

Зазначивши основні здобутки у моделюванні фонологічної системи за попередній період, слід розглянути результати сучасного моделювання. Так, О. В. Кукушкіна, А. А. Поликарпов і Д. В. Хмельов визначили автора тексту, використовуючи математичну модель послідовності букв та граматичних форм як реалізацію ланцюгів Маркова [22]. Модель визначення автора на основі методу опорних векторів використано М. Копелем, Й. Шлером і Ш. Аргамоном [23]. Логічно-структурні моделі як інструмент текстового аналізу та метод дослідження у інших галузях науки широко представлено у роботах В. А. Висоцької, Т. В. Шестакевич, Ю. М. Щербини [24], В. А. Висоцької [25, 26], Т. В. Шестакевич, В. А. Висоцької, Л. В. Чирун, Л. Б. Чирун. [27], С. А. Лупенка [28]. Розглядаючи моделювання семантичних структур при текстовій диференціації у зазначених роботах, слід зауважити, що семантичні структури є абстрактними конструктами реалізація яких залежить від контексту. Тому орієнтація на семантику значною мірою передбачає широке зіставлення конкретних контекстуальних реалізацій. Аналіз статистичних, математичних та лінгвістичних моделей вищезазначених дослідників дає змогу ствердити, що моделі-деталі виокремлені за певною метою дослідника, які репрезентують підсистеми загальної моделі мовної системи,

розроблено без застосування методу комплексного аналізу диференціації фоностатистичних структур стилів та багатофакторного методу визначення ступеню дії факторів стилю, підстилю та авторської манери викладу, і тому рівень достовірності не є достатнім.

Статистичні моделі диференціації стилів української мови розроблено В. І. Перебийніс [17]. Наведемо алгоритм побудови статистичних моделей диференціації стилів української мови на фонологічному рівні у дослідженні В. І. Перебийніс:

- 1) перевірка за критерієм згоди Пірсона, що частоти груп приголосних фонем підлягають закону нормального розподілу;
- 2) диференціація стилів за критерієм Стьюдента;
- 3) лінгвістична інтерпретація статистичних даних.

Для розв'язання завдання підвищення достовірності визначення диференційних ознак стилів, підстилів, та текстів різних авторів необхідно вдосконалити статистичну модель стилевої, підстилевої, та авторської диференціації текстів за методом гіпотез і ранжування.

Для розв'язання завдання підвищення рівня автоматизації диференціації стилів необхідно вдосконалити статистичну модель визначення стилерозрізняльної здатності груп приголосних фонем.

1.4 Диференціація текстів у математичній лінгвістиці

Важливість застосування математичних методів у наукових дослідженнях зазначив вчений світової слави Лорд Кельвін у 1887 році. На його думку: ... якщо ви можете виміряти те, про що ви говорите, і виразити це в числах, тоді ви вже дещо знаєте про це, але якщо ви не можете виміряти те, про що ви говорите, і виразити це в числах, тоді ваше знання недостатнє і незадовільне ... [29].

Подальший розвиток теорії ймовірностей представлений дослідженням В. Я. Буняковського (1846) став фундаментом для виникнення математичної лінгвістики [30]. Значний внесок у розвиток цього напрямку зроблено

Дж. К. Ципфом, який вважає відносну частоту визначальним фактором мовних змін [31]. На думку дослідника: "Вища частотність певної фонемі пояснюється вищим ступенем легкості реалізації цієї фонемі, а нижча – нижчим ступенем легкості її реалізації." [32]. Варто зазначити, що закон Ципфа є актуальним у сучасних дослідженнях. Так, К. Неофіту застосовує цей закон у вивченні частотних характеристик англійської мови у порівнянні з іншими мовами [33].

У межах напряму математичної лінгвістики розвивається лінгвістична статистика. Один із найвидатніших представників напряму математичної лінгвістики Г. Хердан представляє лінгвістичну статистику як квантитативний виклад теорії структурної лінгвістики де Соссюра. Він вважає статистику глибинною філософією мови, яка відображає глибинні протиріччя мови [34].

Основні засади сучасної математичної лінгвістики докладно висвітлено у роботах К. Б. Бектаєва, Р. Г. Піотровського [35, 36], Г. Альтмана, Р. Колера, Р. Г. Піотровського [37]. Подальший розвиток напрям отримав у дослідженнях, в яких окреслено перспективи розвитку математичної лінгвістики із застосуванням інформаційних технологій. Значний внесок у розвиток комп'ютерної лінгвістики в розділі лексикографії зроблено у працях В. А. Широкова, І. В. Остапової, М. В. Надутенко та Ю. І. Вербиненко [38]. Застосування інформаційних технологій широко представлено у дослідженнях Є. А. Карпіловської [39], Н. П. Дарчук [40], О. О. Селіванової [41], А. Корнаї [42], С. Гріса [43], В. А. Висоцької, В. В. Пасічника, Ю. М. Щербини, Т. В. Шестакевич [44], В. С. Перебийніс [45] та інших дослідників.

Доречно виокремити праці, в яких досліджується фонологічний рівень статистичними методами. Основи фонологічної статистики викладено у роботі Д. М. Сегала [46] на матеріалі польської мови. Дослідник зазначає, що статистична однорідність тексту є необхідною при його статистичному обстеженні. Статистичні характеристики фонем китайської мови проаналізовано у роботі Й. Денга [47].

Основним нерозв'язаним завданням попередніх досліджень фонологічного рівня є відсутність методу комплексного аналізу диференціації статистичних структур стилів, який дає змогу підвищити достовірність одержаних результатів.

На сучасному етапі розвитку науково-технічного прогресу дані про частоту вживання мовних одиниць використовуються у системах стенографії в криптографії, у судостилеметрії, при встановленні авторства тексту. Сучасні інформаційні технології (ІТ) широко використовують частотні характеристики для авторської атрибуції тексту на різних рівнях мови. Так, на лексичному рівні тексти розмежовано М. Копелем, Й. Шлером, Ш. Аргамоном, Я. Вінтером, М. Бухейлдом, Дж. Ганацією за частотами службових слів [23, 48 – 51]. Однак, лексичний рівень є відкритою системою. Кількість елементів не є сталою. Система поповнюється новими словами (неологізмами), а рідко вживані слова стають архаїзмами. Авторський стиль відображає змінні процеси лексичної системи. Тому достовірність встановлення авторства на лексичному рівні є нижчою, ніж на фонологічному рівні.

Синтаксичний аналіз українських текстів проведено О. В. Бісікало і В. А. Висоцькою [52]. В. В. Литвин, В. А. Висоцька, І. О. Бобик, П. Я. Пукач і Д. І. Угрин провели граматичний аналіз текстових документів для їх класифікації та визначення авторського стилю застосувавши алгометричну алгебру [53, 54]. У цьому зв'язку варто зазначити, що граматичні структури, які є абстрактними ідеальними сутностями, забезпечують відображення мовленнєвого процесу і уможливають визначення диференційних ознак авторського стилю у конкретному контексті.

Проведений аналіз предметної області дає змогу стверджувати, що у більшості випадків встановлення авторства тексти розмежовуються на лексичному та синтаксичному рівнях мови. Однак фонологічний рівень вирізняється з-поміж інших рівнів мови незмінною кількістю елементів та строгішою структурою. Його легше формалізувати. Тому доцільно встановлювати авторство тексту на фонологічному рівні мови. Отже розроблення методів, моделей та програмних засобів диференціації фоностатистичних структур (статистичних структур на фонологічному рівні – розподілу частот груп приголосних фонем) функціональних стилів англійської мови є актуальним завданням сьогодення.

У попередніх дослідженнях не застосовується поєднання критеріїв Колмогорова-Смірнова, χ^2 -квадрат і Стюдента для диференціації текстів [55 – 57].

При встановленні диференційних ознак авторського стилю статистичними методами, не використано схему стиль→підстиль→автор, яка полегшує визначення статистичних параметрів авторської манери викладу у текстах різної тематики [58 – 61]. Проведений аналіз літературних джерел показав, що завдання підвищення достовірності диференціації текстів не є розв'язаним повною мірою. Для розв'язання завдання необхідно здійснювати авторську атрибуцію на фонологічному рівні, застосувати те поєднання статистичних методів, яке є найефективнішим для одержання достовірних результатів на даному рівні мови та лінгвістичному матеріалі.

1.5 Диференціація текстів у статистичній стилістиці та стилеметрії

З усіх галузей мовознавства стилістика першою скористалась методом статистичного аналізу, тому що цей аналіз дає змогу не тільки виявити відхилення від норми, але й виміряти його. Перші спроби застосувати статистику в художньому стилі зроблено у дослідженні Г. Юла [62], у якому дослідник відстежує залежність стилевих характеристик від довжини речення. Добре відомими є частотні словники поезії, укладені Ж. Майлс [63].

На важливості квантитативного підходу до явищ стилістики неодноразово наголошували багато лінгвістів. Так, наприклад, П. Гіро зазначає, що крива Ципфа дозволяє висунути низку гіпотез про природу мови та її функціонування, а її видозміна є стилістичним критерієм при вивченні конкретного твору і конкретного автора [64, 65]. М. Кожина вважає частоту вживання строго мотивованою та визнає відмінність її величин для кожного із стилів. Тому, на її думку, необхідно вивчати статистичні структури функціональних стилів та широко їх співставляти. При цьому підкреслюється тісний взаємозв'язок між якісною (комплекс засобів та характер їх функціонування) і кількісною (частотність вживання мовних одиниць) сторонами структури певного стилю, тому що обидві визначаються певною ціллю спілкування і залежністю від конкретних екстралінгвістичних факторів [66].

Стиль є суспільно необхідним, історично сформованим різновидом літературної мови (її функціональною підсистемою), що обслуговує певну сферу суспільної діяльності мовців і відповідно до цього має свої особливості добору і використання мовних одиниць [67]. Важливе значення у наведеному визначенні має наголошення на особливостях добору мовних одиниць, які залежать від конкретної сфери діяльності. Однак, слід пам'ятати, що кожен стиль не утворює окремої замкненої системи. Часткова відмінність стилів означає наявність спільних елементів, що пояснюється взаємопроникненням елементів стилів.

Взаємопроникнення елементів стилів доречно розглядати на рівнях мови і мовлення. Функціональні стилі на рівні мови – рівні дещо абстрактному ідеальному, відзначаються чіткішим розмежуванням, а на мовленнєвому рівні – рівні численних реалізацій мовних явищ, вони, природно, мають більш розмиті, невизначені границі. До екстралінгвальних чинників, які впливають на стилістичну забарвленість тексту, можна віднести випадкову або постійну дію авторської манери викладу. Якщо авторська манера викладу може деформувати у певною мірою межі, якими визначається певний стиль, то для визначення стилістичної приналежності тексту необхідно розкрити її особливості. Пошук параметрів, які б сприяли усуненню різнобою у виокремленні функціональних стилів, здійснювався багатьма дослідниками. Так, Л. М. Гриднева вивчає закономірності розподілу голосних, приголосних і пропусків у функціональних стилях українського мовлення [68]. Вагомим є імовірісно-статистичне вивчення стилістичної диференціації мови Б. Н. Головіна [69] та квантитативно-системний аналіз лексики проведений Ю. А. Тулдавою [70]. Найбільшу повноту, різнобічність і систематизованість статистичне обстеження текстів отримує у монографії В. С. Перебийніс, в якій зроблено спробу встановити статистичні параметри стилів української мови [21]. Отримані В. С. Перебийніс результати переконливо доводять доцільність застосування методів математичної статистики, які дають змогу повніше розкрити якісну природу певного стилістичного компонента досліджуваного твору, пізнати внутрішню структуру стилю, точніше виміряти розходження між стилями чи авторами.

На сучасному етапі застосування методів математичної статистики у дослідженні авторського стилю вагомим є дослідження В. Гульєн Ньето, Ч Варгас-Сьєрра, М. Пардіфіно Жуан та А. Суарез Куето, які здійснюють авторську атрибуцію у судочинстві. Авторські стилі розмежовуються за різницею відносних частот артиклів, займенників, прийменників та допоміжних дієслів. Застосовано статистичний критерій χ^2 -квадрат. Результати отримано з довірчою ймовірністю 90%, [71]. Г. Байєн і Г. ван Гальтерен використовують дискримінантний аналіз для розрізнення текстів різних авторів. Аналізується відмінність частот службових слів з довірчою ймовірністю 81,5% [72]. Г. Раджу, Ч. Садхві, П. Теджасвіні та Й. Муніка використовують опорно-векторний метод для встановлення авторства тексту, аналізуючи відмінність частот літер та службових слів з довірчою ймовірністю – 97,22% [73]. Т. Петменсон застосовує статистичний критерій Стюдента та коефіцієнти рангової кореляції, розмежовуючи перші слова речень та знаки пунктуації у зіставлених авторських стилях з довірчою ймовірністю 66,7% [74]. І. А. Волошиновська використовує для встановлення авторства тексту метод рангово-частотного розподілу слів [75]. Зазначивши попередній доробок дослідників, слід зауважити, що рівень достовірності отриманих результатів не є достатнім, бо тексти не диференціюються потужним непараметричним критерієм Колмогорова-Смірнова. Критерій Колмогорова-Смірнова використано О. Гранічиним в авторській атрибуції, при розмежуванні літер та сполук з двох-трьох елементів [76]. Однак, з огляду 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 зв'язку логічним є пошук поєднання найдієвіших статистичних методів. Поєднання потужних параметричних і непараметричних критеріїв диференціації стилів забезпечує підвищення достовірності результатів.

Оскільки ідея поєднання статистичних методів є відсутньою в існуючих 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. Аналізуюючи рівні мови з погляду ступеню впорядкованості елементів та змінності їх кількості, оптимальним рівнем можна вважати фонологічний. Впорядкованість елементів синтаксичного рівня залежить від структури речення, яке може мати просту, складну та ускладнену структуру. Переваги фонологічного рівня над іншими рівнями мови показано на рис. 1.5.1. 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 на фонологічному рівні) структур стилів. У дослідженні фоностатистична структура стилю – розподіл частот груп приголосних фонем визначених для певного стилю.

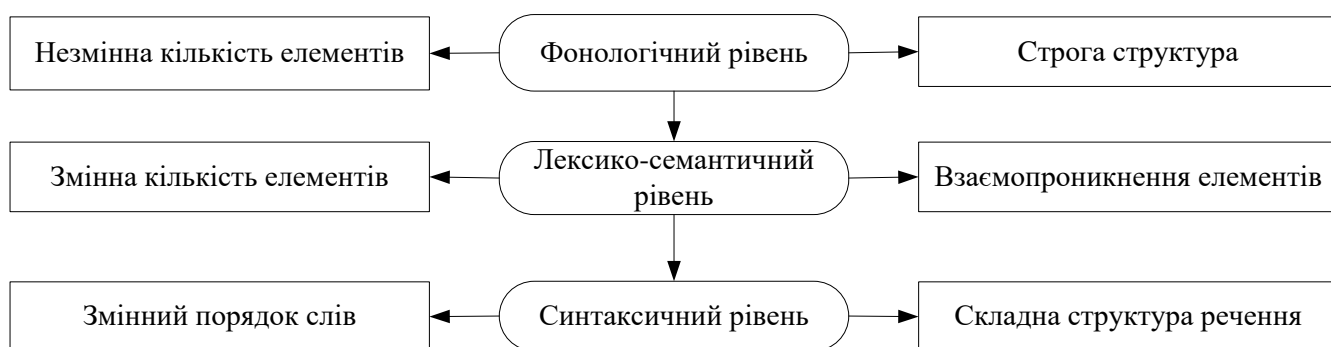


Рисунок 1.5.1 – Специфіка фонологічного рівня

Проведений аналіз використуваних методів показав, що в сучасних дослідженнях відсутній багатофакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу, відповідно до якого, диференціація стилів здійснюється за схемою, яка зображена на рис. 1.5.2. Ідея методу полягає у визначенні загальної стилевої маркованості досліджуваного тексту, яка є рівнодіючою факторів стилю, підстилю та авторської манери викладу. Визначення загальної стилевої маркованості досліджуваного тексту певного автора дає суттєву інформацію про характерні особливості авторського стилю, які є диференційними ознаками, за якими даний авторський стиль відрізняється від інших авторських стилів. Істотні відмінності між авторськими стилями встановлюються за різницею значень загальної стилевої маркованості зіставляваних попарно текстів. Метод дає

змогу підвищити достовірність авторської диференціації текстів. Тому актуальною складовою для досягнення мети дисертаційного дослідження є розроблення багатofакторного методу.

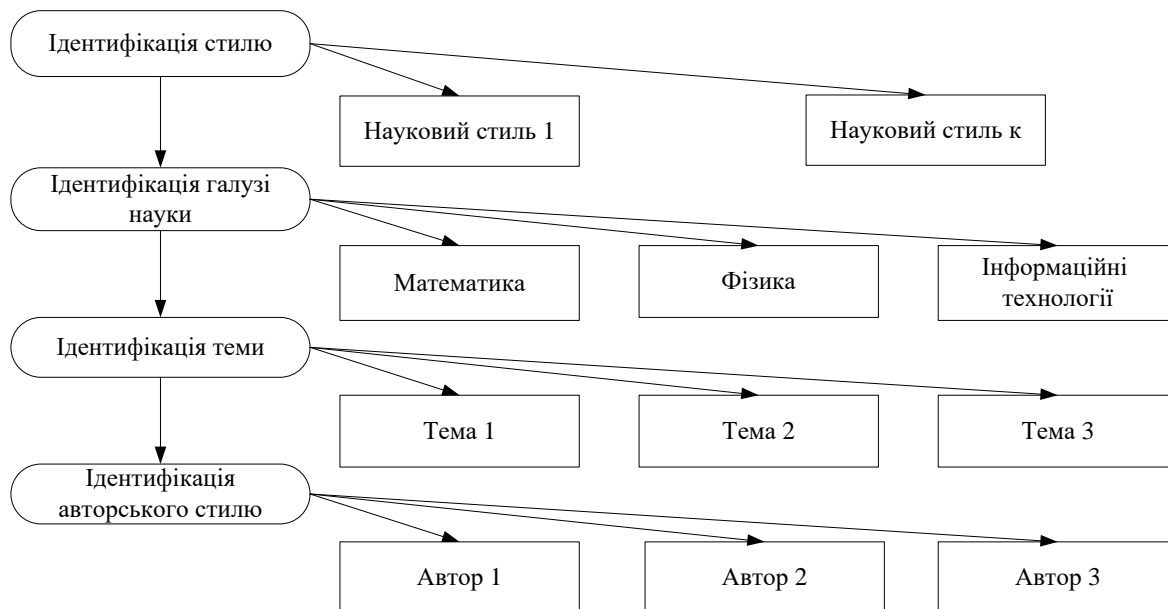


Рисунок 1.5.2 – Схема диференціації стилів

В останні роки статистична стилістика представлена дослідженням Е. Вілсон, в якому аналізуються статистичні характеристики елементів просодики [77].

Отже, застосовуючи методи математичної статистики, можна отримати статистичні параметри, які шляхом їх стилістичної інтерпретації дають можливість чіткіше визначити особливості кожного із стилів.

1.6 Аналіз програмних засобів диференціації стилів

Аналіз засобів диференціації текстів показав, що ефективними є програмні системи від NeoNeuro, Unplag, Quora, EVL, JGAAP, JGAAP in version 3.1 [78 – 83].

Для прикладу подаємо алгоритм функціонування програми від NeoNeuro:

- 1) завантажити текст, авторство якого потрібно встановити;
- 2) завантажити тексти авторів, із стилями яких зіставляється досліджуваний текст;

3) проаналізувати дані розміщені у порядку зростання ймовірності приналежності певному автору. Цифра "1" позначає автора, якому належить досліджуваний текст [78].

Схожим є алгоритм ще однієї online програми:

- 1) скопіювати і вставити досліджуваний текст першого автора в "Author 1" box;
- 2) скопіювати і вставити досліджуваний текст другого автора в "Author 2" box;
- 3) скопіювати і вставити досліджуваний текст невідомого автора в "Unknown author" box;
- 4) натиснути "submit button" для одержання даних про приналежність досліджуваного тексту першому або другому автору [81].

Встановлення авторства тексту пов'язане з перевіркою тексту на плагіат. Ефективність Unplag програми перевірки робіт на плагіат підтверджена використанням її в більш, ніж 45 українських університетах. Програму розроблено в Україні. Сервіс активно використовують в США, Іспанії, Бельгії та інших країнах.

Використовується програма The Java Graphical Authorship Attribution Program (JGAAP) розроблена професором Патріком Джолою та його співробітниками в лабораторії EVL Labs, Дюкейнського університету, США. Для встановлення авторства тексту використовується метод лінійної регресії. Дослідження проведено на лексичному рівні [82].

Зазначивши ефективність розглянутих програм, слід наголосити, що 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 програмої системи авторської, підстилевої, і стилевої атрибуції тексту на фонологічному рівні. Застосоване попередніми дослідниками програмне забезпечення не базується на використанні поєднання статистичних методів та диференціації текстів за значенням загальної стилевої маркованості і тому не забезпечує одержання результатів належного рівня достовірності.

1.7 Висновки до розділу 1

Зазначивши вагомий доробок попередніх досліджень, слід наголосити, що застосовані дослідниками методи дали змогу отримати різні значення достовірності авторської атрибуції тексту. 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 більшої достовірності, її можна отримати застосувавши поєднання ефективних статистичних методів, за якими встановлюються істотні відмінності на тому самому матеріалі. Крім цього, сучасні інформаційні технології дають змогу опрацьовувати великий обсяг інформації, що забезпечує проведення статистичного обстеження текстів різних стилів, підстилів та авторів. Тому, з огляду на сучасний стан досліджень, актуальним є наукове завдання підвищення достовірності диференціації стилів на основі розроблених методів, моделей і програмних засобів. У цьому зв'язку у дисертаційному дослідженні необхідно виконати такі завдання:

- 1) розробити метод комплексного аналізу диференціації фоностатистичних структур стилів для підвищення достовірності диференціації стилів;
- 2) розробити багатофакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу для встановлення авторської, підстилевої та стилевої атрибуції досліджуваного тексту;
- 3) удосконалити статистичні моделі для одержання вихідних параметрів на основі результатів статистичного обстеження текстів;
- 4) розробити програмні засоби для диференціації фоностатистичних структур стилів англійської мови.

РОЗДІЛ 2. РОЗРОБЛЕННЯ СТАТИСТИЧНИХ МЕТОДІВ ДЛЯ ДИФЕРЕНЦІАЦІЇ ФОНОСТАТИСТИЧНИХ СТРУКТУР СТИЛІВ АНГЛІЙСЬКОЇ МОВИ

2.1 Розроблення методу комплексного аналізу для диференціації фоностатистичних структур стилів англійської мови

В математичній лінгвістиці особливе місце займає теорія ймовірностей і її прикладний розділ – математична статистика. Математична статистика є потужним апаратом, який дозволяє на основі даних вибірки обмеженого обсягу робити відповідні висновки про генеральну сукупність, тобто про всю досліджувану систему [84]. Для того, щоб зробити ці висновки, треба вибрати певний критерій диференціації стилів. Вибір критерію розмежування стилів залежить від методу, який застосовується [85]. Вперше розроблений метод комплексного аналізу диференціації фоностатистичних структур стилів англійської мови ґрунтується на поєднанні статистичних методів гіпотез, ранжування і визначення відстаней між стилями, що дає змогу підвищити достовірність диференціації стилів. Підвищення достовірності розмежування текстів різних стилів, підстилів та авторів забезпечується отриманням однакового результату за кількома методами, які є складовими методу комплексного аналізу. Найвищого рівня достовірності можна досягнути, отримавши істотні відмінності за трьома методами, середнього – за двома методами, найнижчого – за одним методом.

На рис. 2.1 зображено блок-схему алгоритму методу комплексного аналізу.

У дослідженні середні значення частот груп приголосних фонем, як критерій диференціації стилів, повинні давати наступну інформацію:

- в якій мірі ці величини, одержані з вибірки обмеженого обсягу, характеризують стиль;

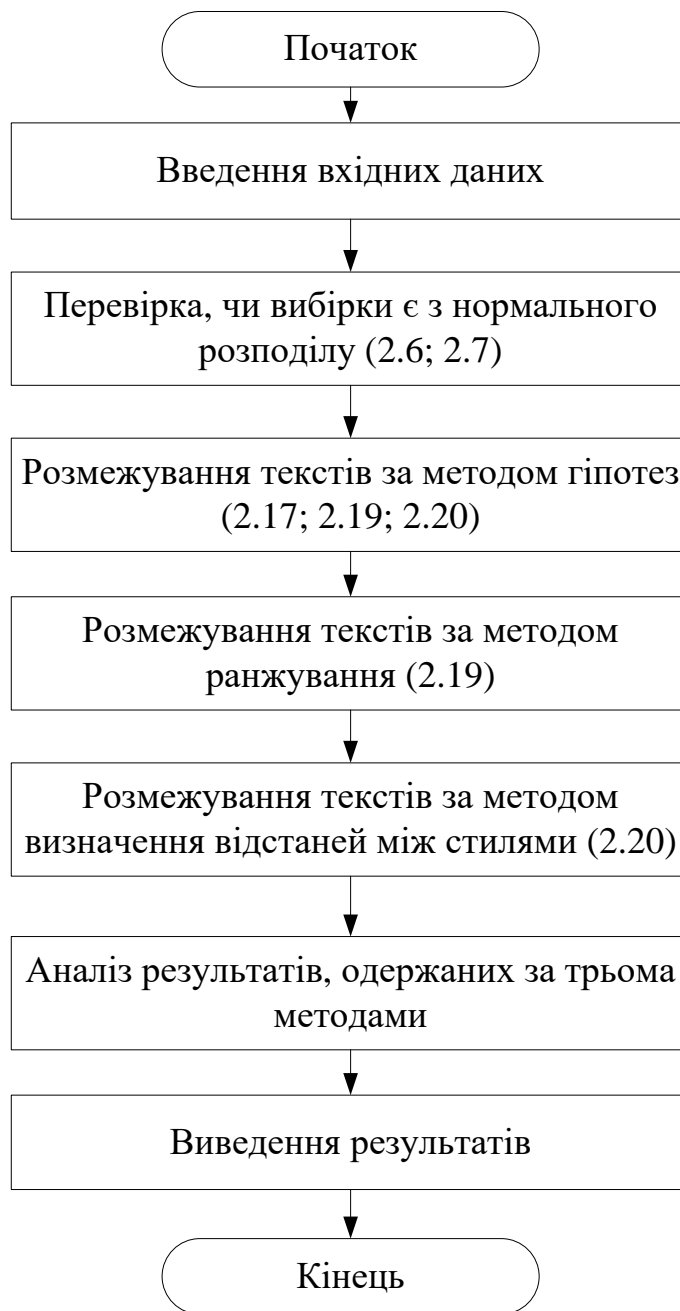


Рисунок 2.1 – Блок-схема алгоритму методу комплексного аналізу диференціації фоностатистичних структур стилів англійської мови

- чи можуть вони бути критерієм відмінності стилів, чи характеризують англійську мову в цілому.

Щоб відповісти на поставлені питання, вводиться рівень значущості – ймовірність появи певного значення різниці середніх значень частот порівнюваних

систем, якою можна нехтувати [86]. Якщо в результаті досліджень одержуємо величину, ймовірність появи якої дорівнює або менша від рівня значущості, то порівнювані системи відрізняються істотно. У протилежному випадку різниця між цими системами неістотна (в плані вибраного критерія) і зумовлена звичайним статистичним розкидом [87].

З точки зору засад теорії ймовірностей розподіл значень досліджуваної величини дає практично всю інформацію про цю величину [88, 89]. Тому цікаво визначити розподіл частоти досліджуваних груп приголосних фонем. Можна вважати апіорі, що це є нормальний розподіл. В науці нормальний розподіл відіграє особливу роль. Багато характеристик строгих фізичних моделей, які віддзеркалюють реальний світ, підлягають нормальному розподілу [90 – 92]. Мова не є винятком. Вона також є системно-структурним утворенням. Отже, можна припустити, що частота кожної з досліджуваних груп приголосних фонем підлягає нормальному розподілу. Для перевірки цього твердження у дослідженні використовується вибірка обсягом 714 тисяч фонем (714 порцій по тисячу фонем). Вона складається із 14 менших вибірок по 51 тисячі фонем в кожній, взятих відповідно з текстів художнього стилю (поезії і художньої прози Дж. Г. Байрона, поезії Т. Мура, драми Б. Шоу), розмовного, газетного, публіцистичного і наукового стилів.

Для розмежування зазначених стилів, розроблено метод комплексного аналізу фоностатистичних структур стилів англійської мови, який ґрунтується на поєднанні методів гіпотез, ранжування і визначення відстаней між стилями.

2.2 Використання методу гіпотез для диференціації фоностатистичних структур стилів англійської мови

Для диференціації стилів використано метод гіпотез, складовими якого є критерій Стюдента, критерій Колмогорова-Смірнова та критерій хі-квадрат.

Блок-схема алгоритму методу гіпотез зображена рис. 2.2.

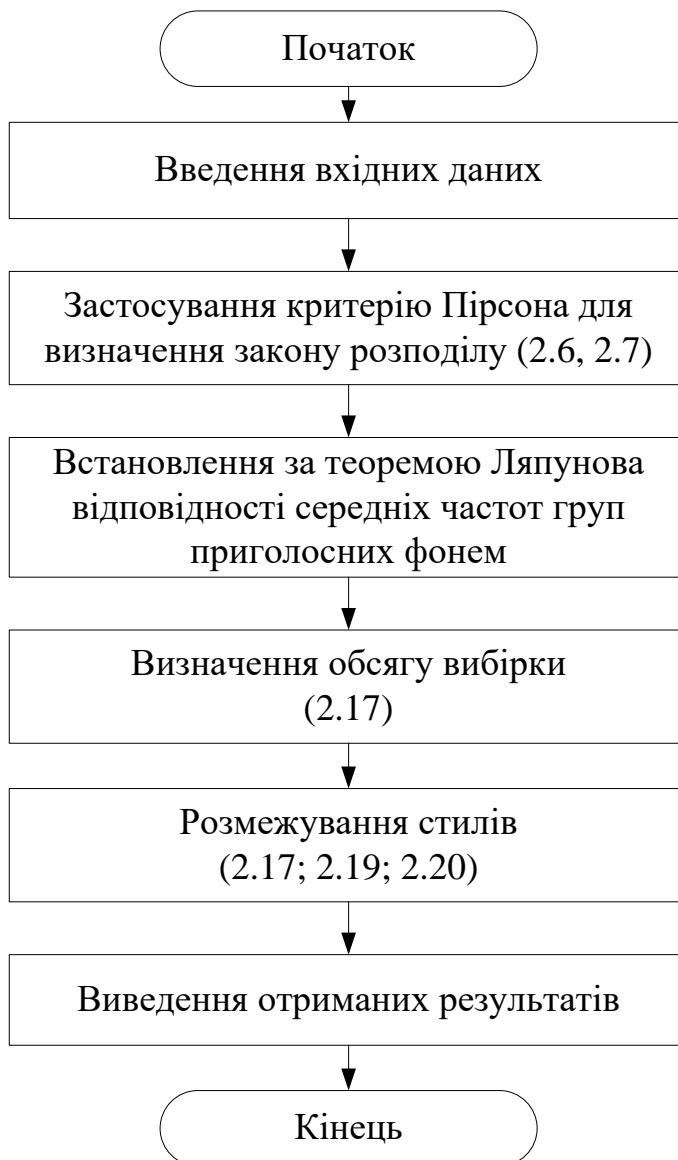


Рисунок 2.2 – Блок-схема алгоритму методу гіпотез

Диференціація текстів за критерієм Стюдента передбачає відповідність частот груп приголосних фонем закону нормального розподілу. Тому висуваємо нульову гіпотезу H_0 – одержані емпіричні дані відповідають закону нормального розподілу [93].

Для перевірки цієї гіпотези необхідна деяка статистична характеристика, яка оцінює розбіжність емпіричного і теоретичного розподілів. За критерій згоди

використовуємо критерій χ^2 Пірсона [93]. Цей критерій є добрий тим, що його статистика не залежить від закону розподілу досліджуваної випадкової величини.

Розподіл Пірсона визначає при заданій кількості ступенів вільності ν імовірність $\rho(x^2)$ одержати значення x^2 , яке є більше від деякого x_0^2 .

Для χ^2 -розподілу Пірсона побудована таблиця, яка дає змогу при заданій кількості ступенів вільності ν і рівні значущості g визначити x_0^2 [93, 33]. При здійсненні конкретних розрахунків замість x_0^2 звичайно пишуть $x_{g;\nu}^2$.

Якщо в результаті проведених досліджень одержано $x^2 > x_{g;\nu}^2$, то гіпотеза відхиляється. У протилежному випадку вона приймається. Іншими словами, це означає наступне: якщо в результаті обчислень одержано значення величини x^2 , ймовірність появи якого $\rho(x^2)$ при кількості ступенів вільності ν є менша за рівень значущості g , емпіричний розподіл частоти досліджуваної групи фонем не є нормальним. Якщо ж $\rho(x^2) \geq g$, то частоти відповідають закону нормального розподілу. Чим більше $\rho(x^2)$, тим краще емпіричний розподіл наближається до нормального розподілу. При встановленні істинності нульової гіпотези H_0 використовується схема обчислень, запропонована К. Б. Бектаєвим і Р. Г. Піотровським [33]. Значення частот X кожної групи фонем займають певну область числової осі. Цю область розбиваємо на m інтервалів шириною ΔX . Кожний інтервал має нижню границю α_i і верхню β_i , причому $\beta_i = \alpha_{i+1}$.

В кожний інтервал потрапляє n_i частот. Середня частота визначається за формулою:

$$\bar{X} = \frac{\sum_{i=1}^m x_i n_i}{N}, \quad (2.1)$$

де x_i – середина інтервалу Δx , $N = 714$ – кількість порцій вибірки.

Варто зауважити, що в деяких випадках $N = 713$, тому що в одній з порцій частота істотно відрізняється від частот решти порцій, і згідно з загальними засадами теорії ймовірностей повинна бути відкинута. Крім того, у деяких випадках

область значень величини X зліва і справа збільшена на 0,5. Це практично не впливає на обчислення, але усуває невизначеність у потраплянні певних частот в той чи інший інтервал.

На наступному етапі визначаємо теоретичний нормальний розподіл частоти досліджуваної групи фонем. Ймовірність потрапляння випадкової величини X в певний інтервал $\Delta x = \beta_i - \alpha_i$ дорівнює:

$$\rho_i = (\beta_i - \alpha_i) f(x), \quad (2.2)$$

де $f(x)$ – густина нормального розподілу:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (2.3)$$

де μ і σ^2 відповідно середнє значення і дисперсія частоти генеральної вибірки. Ці величини є невідомими. Замінюємо їх на \bar{x} і S^2 , де S^2 – незміщена оцінка дисперсії:

$$S^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 n_i}{N-1}. \quad (2.4)$$

Використовуючи вищесказане, одержуємо наступний вираз для теоретичного значення частоти:

$$n_i = \frac{N\Delta x}{S} \varphi(z), \quad (2.5)$$

де $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$; $z = \frac{x - \bar{x}}{S}$.

Функція $\varphi(z)$ є табульованою [93, 33].

Далі перевіряється істинність гіпотези H_0 . Значення критерію χ^2 обчислюється за формулою:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - n_i')^2}{n_i'} \quad (2.6)$$

Використовується 5%-й рівень значущості, тобто $g = 0,05$. Для визначення кількості ступенів вільності ν необхідно щоб частота нормального розподілу, встановленого на основі емпіричного розподілу, була підпорядкована трьом зв'язкам (сума спостережених частот фіксована; теоретичні частоти повинні давати середню, рівну \bar{x} емпіричних частот; дисперсії теоретичного і емпіричного розподілів повинні бути рівні) [93, 33].

Варто зауважити, що вираз (2.6) одержано за умови, що біноміальний розподіл може бути зведений до нормального. Це можливо тоді, коли n_i інтервалів не будуть надто малими. Тому при використанні критерія χ^2 , n_i' розрядів, якщо вони малі (менші п'яти), необхідно з'єднувати між собою [93, 33].

Це враховано у формулі (2.6): у результаті з'єднання інтервалів з малими n_i , їх кількість зменшується ($k < m$). Враховуючи всі міркування, одержуємо $\nu = k - 3$.

Також для перевірки гіпотези H_0 використовується спрощений критерій Романовського, який дає змогу визначити близькість емпіричного і теоретичного розподілів [93, 33]. Згідно з цим критерієм, якщо

$$\frac{|x^2 - \nu|}{\sqrt{2\nu}} \geq 3, \quad (2.7)$$

то розходження між емпіричним і теоретичним розподілами треба вважати істотним; якщо ж

$$\frac{|x^2 - \nu|}{\sqrt{2\nu}} < 3, \quad (2.8)$$

то це розходження треба вважати випадковим.

Результати перевірки відповідності частот закону нормального розподілу представлено у додатку Б у восьми таблицях Б.1 – Б.8 і восьми графіках В.1 – В.8. На графіках суцільна крива – теоретичні дані, окремі точки відповідають емпіричним даним. За даними таблиць можна зробити висновок, що для всіх досліджуваних груп фонем задовільняється 5%-ий рівень значущості, тобто

$\chi^2 < \chi_{0,05; \nu}^2$. У більшості випадків показники набагато кращі: імовірність одержати обчислену величину χ^2 значна: $\rho(\chi^2)$ знаходиться в інтервалі від десяти до сорока п'яти відсотків. Як критерій Пірсона, так і спрощений критерій Романовського підтверджують істинність гіпотези H_0 : емпіричні розподіли для всіх восьми груп приголосних фонем підлягають нормальному розподілу. Отже, частоти груп приголосних фонем розподіляються згідно з рівнянням (2.3). Величина σ є середньоквадратичним відхиленням нормального розподілу. Згідно із засадами теорії ймовірностей, при 5%-ому рівні значущості математичне сподівання частот груп приголосних фонем з ймовірністю 0,95 знаходиться у межах:

$$\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}} < X < \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}, \quad (2.9)$$

де \bar{X} – середнє значення частоти групи фонем.

Зважаючи на те, що у дослідженні критерієм диференціації є середні значення частот груп приголосних фонем, необхідна інформація про ці величини. З цією метою використовуємо теорему Ляпунова [93, 94]. З теореми Ляпунова випливає, що розподіл середнього значення випадкової величини X є нормальним із середнім значенням, яке дорівнює середньому значенню цієї випадкової величини, і середньоквадратичним відхиленням σ / \sqrt{n} , де n – кількість порцій вибірки.

Істотним в подальшому дослідженні є наступне. Параметри вибірки обмеженого обсягу можна вважати точковими оцінками параметрів генеральної сукупності [94]. Необхідно на основі цих даних (тобто на основі тексту, з яким ми працюємо) одержати інформацію про генеральну сукупність – певний стиль англійської мови. З цією метою використано статистику:

$$t = \frac{\bar{x} - \bar{x}_0}{S} \sqrt{n}, \quad (2.10)$$

де \bar{x} – середнє значення частоти груп фонем, одержане з вибірки, з якою ми працюємо; n – кількість порцій цієї вибірки; \bar{x}_0 – середнє значення частоти групи

приголосних фонем генеральної сукупності; S^2 – незміщена оцінка дисперсії. Статистика t підлягає розподілу Стьюдента [93].

Статистика t дає змогу встановити границі середнього значення \bar{X}_0 . Зокрема, при 5%-ому рівні значущості, середнє значення частоти груп приголосних фонем генеральної сукупності знаходиться з довірчою ймовірністю 0,95 у таких межах:

$$\bar{X} - t_{0,05} \frac{S}{\sqrt{n}} < \bar{X}_0 < \bar{X} + t_{0,05} \frac{S}{\sqrt{n}}. \quad (2.11)$$

Значення $t_{0,05}$ знаходимо з таблиці [93]. Ця величина відповідає 5% рівню значущості і числу ступенів вільності $\nu = n - 1$.

Для наочності розглянемо границі величини $\bar{X}_0 - \bar{X}$ – значення різниці середніх частот генеральної сукупності та даної вибірки. З (2.11) одержуємо:

$$-t_{0,05} \frac{S}{\sqrt{n}} < \bar{X}_0 - \bar{X} < t_{0,05} \frac{S}{\sqrt{n}}. \quad (2.12)$$

Введемо наступні позначення: $\Delta = k_n S$; $k_n = \frac{t_n}{\sqrt{n}}$, де t_n є значення величини t для заданої кількості порцій n і заданого рівня значущості.

Отже, величина $\bar{X}_0 - \bar{X}$ при вибраному рівні значущості 0,05 з імовірністю 0,95 знаходиться у межах:

$$-\Delta < \bar{X}_0 - \bar{X} < \Delta. \quad (2.13)$$

Досвід фоностатистичних досліджень показує, що при збільшенні n величина Δ зменшується. З точки зору засад теорії ймовірностей це очевидний факт: при збільшенні n спостерігається наближення до генеральної сукупності. Отже, при великих n одержуємо практично достовірну інформацію про стиль англійської мови (в плані вибраного критерію диференціації).

Приклад залежності Δ від n для групи середньоязикових фонем наведено в Додатку Д. Кривим 1, 2, 3 відповідає: художня проза Байрона, поезія Байрона, науковий стиль. Точкам на осі абсцис відповідають числа 2; 6; 11; 16; 21; 26; 31.

Все вищевикладене дозволяє розв'язати основне завдання, а саме: показати, як на основі вибраного критерію диференціації можна диференціювати функціональні

стилі англійської мови. Критерієм диференціації вибрано середні значення частот груп приголосних фонем. Це групи губних, передньоязикових, середньоязикових, задньоязикових, носових, сонорних, щілинних і зімкнених.

Об'єктом нашого дослідження є наступні функціональні стилі: науковий, газетний, публіцистичний, розмовний та художній, який охоплює три підстилі: поезію, художню прозу та драму. Авторський стиль досліджується на матеріалі публіцистичного і художнього стилів. Науковий, газетний та художній стилі порівнюються один з одним і з розмовним стилем.

У плані квантитативного підходу дослідження стилі характеризуються набором чисел \bar{x}_r^{-a} . Верхній індекс позначає групу приголосних фонем. В дослідженні використовується вісім груп фонем. Отже, a приймає значення 1,2,...,8. Досліджується п'ять стилів шляхом зіставлення їх між собою і з точкою відліку. Отже, r приймає значення 1,2,...,7.

Для розв'язання поставленого завдання (чи можна за допомогою величин \bar{x}_r^{-a} диференціювати функціональні стилі) висуваємо гіпотезу: одержані величини \bar{x}_r^{-a} характеризують англійську мову в цілому, і тому незбіг їх значень є неістотним, випадковим. У цьому випадку вони характеризують одну спільну сукупність. Перевірка гіпотези полягає у з'ясуванні факту: чи при вибраному рівні значущості 0,05 значення різниць $\bar{x}_1^{-a} - \bar{x}_2^{-a}$; $\bar{x}_2^{-a} - \bar{x}_3^{-a}$ і т.д. є величинами істотними. Слід заважити, що у дослідженні проводились експерименти з різним рівнем значущості: 0,05, 0,03, 0,01. Однак найоптимальнішим рівнем є рівень значущості 0,05 і тому цей рівень вибрано у більшості експериментів.

Розглядаємо дві вибірки з різних текстів (стилів), які характеризуються середніми частотами вибраної групи приголосних фонем, наприклад \bar{x}_1^{-a} і \bar{x}_2^{-a} [93].

Оцінка дисперсії цих величин визначається за формулою:

$$S_1^2 = \frac{\sum_{i=1}^{31} (x_{1i} - \bar{x}_1)^2}{n_1 - 1}. \quad (2.14)$$

Перевірка висунутої гіпотези полягає в оцінці різниць типу $\bar{x}_1^a - \bar{x}_2^a$. Дисперсія останньої величини дорівнює:

$$\sigma_{\bar{x}_1^a - \bar{x}_2^a}^2 = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \frac{n_1 + n_2}{n_1 \cdot n_2} \cdot \sigma^2. \quad (2.15)$$

У виразі (2.15) σ – дисперсія загальної сукупності, n_1 і n_2 – кількість порцій відповідно у першій і другій вибірках. Тому що оцінки S_1^2 і S_2^2 дисперсії σ^2 мають вагу, відповідно $n_1 - 1$ і $n_2 - 1$, повна оцінка дисперсії дорівнює:

$$S^2 = \frac{\sum_{i=1}^{n_1} (x_{1i}^a - \bar{x}_1^a)^2 + \sum_{i=1}^{n_2} (x_{2i}^a - \bar{x}_2^a)^2}{n_1 + n_2 - 2}. \quad (2.16)$$

Далі одержуємо:

$$t = \frac{\bar{x}_1^a - \bar{x}_2^a}{S} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}. \quad (2.17)$$

У дослідженні $n_1 = n_2 = 31$. Звідси замість (2.16) одержуємо таку робочу формулу:

$$t = 3,937 \cdot \frac{\bar{x}_1^a - \bar{x}_2^a}{S}. \quad (2.18)$$

Величина t є табульованою [93]. Таблиця показує ймовірність, з якою з'являється конкретне значення величини t . Чим більше це значення, тим менша ймовірність появи конкретного значення величини t .

Для дисертаційного дослідження поява значення величини t , якій відповідає мала ймовірність, свідчить, що величини \bar{x}_r^a істотно відрізняються одна від одної. У цьому випадку вони є характеристиками стилів, а не англійської мови в цілому. Правомірність цього твердження збільшується із збільшенням t .

При оцінці різниць типу $\bar{x}_1^a - \bar{x}_2^a$ застосовується двосторонній рівень значущості $2Q$. Цей критерій дорівнює ймовірності того, що t за абсолютною величиною дорівнюватиме або буде більше за деяку величину. У більшості експериментів вибрано рівень значущості 0,05. Отже, якщо $2Q < 0,05$, то різниці типу $\bar{x}_1^a - \bar{x}_2^a$ є істотними, і величини \bar{x}_r^a характеризують окремі стилі.

Таким чином, з вищевикладеного випливає, що, працюючи з текстами обмеженого обсягу, на основі вибраного критерію диференціації (середніх значень груп приголосних фонем) одержано інформацію про відмінності між стилями. Дані досліджень представлено у Додатку А. Сім вибірок, які представляють наступні функціональні стилі: розмовний, газетний, науковий і художній (поезія, художня проза і драма) порівнюються попарно. В зв'язку з тим, що дослідження проведено для випадків невизначеної позиції фонемі в слові і для випадків позиції фонемі на початку та в кінці слова, загальна кількість таблиць: $3 \cdot C_7^2 = 63$. Таблиці представлено у Додатку А.

Крім критерію Стюдента, у дисертаційному дослідженні використано критерій Колмогорова-Смірнова та критерій хі-квадрат на матеріалі семи вибірок з художнього та публіцистичного стилів. Підтвердження гіпотези на однорідність текстів за критерієм Колмогорова-Смірнова, який є найпотужнішим критерієм у даному поєднанні критеріїв (рис. 2.3), здійснюється таким чином:

$$\lambda_{n,m} = \sqrt{\frac{nm}{n+m}} D_{n,m} = \sqrt{\frac{nm}{n+m}} \sup_{-\infty < z < \infty} |F_n(z) - F_m(z)|, \quad (2.19)$$

де $D_{n,m} = \sup_{-\infty < z < \infty} |F_n(z) - F_m(z)|$; $F_n(z)$ і $F_m(z)$ – емпіричні функції розподілу, побудовані для двох вибірок n і m , $\lambda_{n,m}$ – значення статистики Колмогорова-Смірнова.

За заданим рівнем значущості α знайдемо за таблицею критичне значення λ_α . Якщо $\lambda_{n,m} \geq \lambda_\alpha$, то гіпотеза H_0 відхиляється, якщо ж $\lambda_{n,m} < \lambda_\alpha$ – то приймається [95].

Підтвердження гіпотези на однорідність текстів за критерієм хі-квадрат здійснюється таким чином:

$$\hat{\chi}_n^2 = \sum_{i=1}^s \sum_{j=1}^k \frac{\left(v_{i,j} - \frac{n_j v_{i.}}{n} \right)^2}{\frac{n_j v_{i.}}{n}}, \quad v_{i.} = \sum_{j=1}^k v_{i,j}, \quad (2.20)$$

де $v_{i,j}$ – кількість елементів, які потрапили в i -ий інтервал j -того тексту, $v_{i..}$ – кількість елементів, які потрапили в i -ий інтервал двох текстів, s – кількість інтервалів, k – кількість текстів, n – кількість порцій в двох текстах, n_j – кількість порцій в одному тексті.

Теорема 2.1 (Гранична теорема) [96, 97]. Якщо $0 < p_i < 1$, $i = 1, 2, \dots, N$, то при $n \rightarrow \infty$ розподіл величини $\hat{\chi}_n^2$ слабо збігається з χ^2 -розподілом з $(N-1)$ ступенями вільності.

Аналогічною до теореми 2.1 є гранична теорема 2.2, яка застосовується для визначення критичної границі.

Теорема 2.2: $\hat{\chi}_n^2 \xrightarrow[n \rightarrow \infty]{ct} \chi^2((s-1)(k-1))$.

У таблиці χ^2 -розподілу за заданим рівнем значущості α та кількістю ступенів вільності $(s-1)(k-1)$ знаходимо число $\chi_{1-\alpha, (s-1)(k-1)}^2$ таке, що $P\{\chi^2(s-1)(k-1) \geq \chi_{1-\alpha, (s-1)(k-1)}^2\} = \alpha$. Якщо значення статистики критерію $\hat{\chi}_n^2 \geq \chi_{1-\alpha, (s-1)(k-1)}^2$, то гіпотеза H_0 відхиляється [99 – 103].

Визначену за одержаними даними потужність застосованих критеріїв зазначено на рис. 2.3. Результати отримані за критерієм Стюдента, критерієм Колмогорова-Смірнова та критерієм хі-квадрат представлено у підрозділі 4.6.

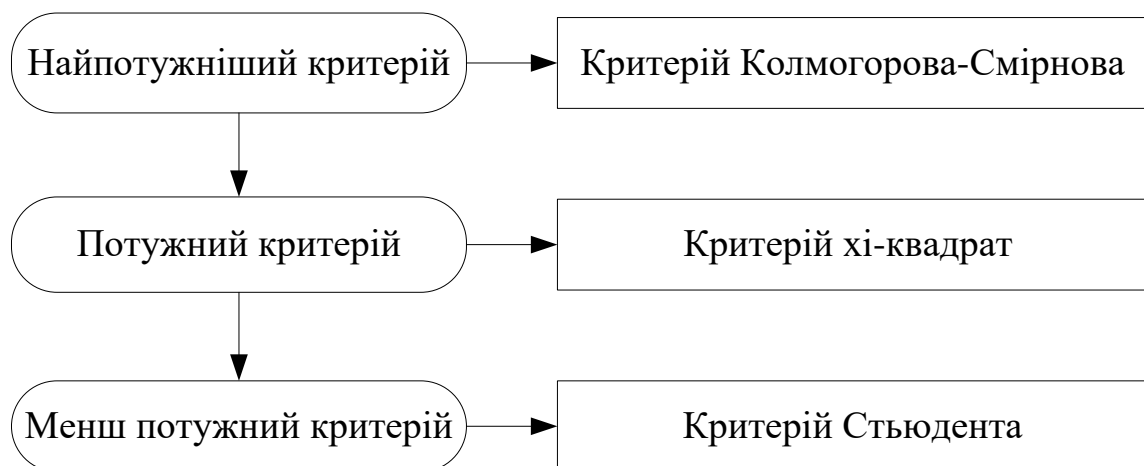


Рисунок 2.3 – Схема рівня потужності критеріїв диференціації стилів

Отже, підвищення достовірності диференціації стилів, підстилів та текстів різних авторів забезпечується застосуванням п'яти методів, одним з яких є непараметричний метод, потужність якого прямує до 1 (критерій Колмогорова-Смірнова) [100]. Результати отримано з довірчою ймовірністю 0,95.

2.3 Розроблення методу ранжування для диференціації фоностатистичних структур стилів англійської мови

Для диференціації функціональних стилів на фонологічному рівні використано метод ранжування емпіричних даних.

Розроблена блок-схема алгоритму методу ранжування зображена на рис. 2.4.

Емпіричними даними у дослідженні є середні значення частот груп приголосних фонем функціональних стилів – величини \bar{X}_r^a . Ранжування полягає у впорядкуванні емпіричних даних, тобто у побудові ряду величин \bar{X}_r^a . У дисертаційному дослідженні використовуються спадні ряди. Номер місця у ряді, яке займає певна величина \bar{X}_r^a , називається рангом цієї величини або ранговим показником. Рангові показники надаються середнім частотам груп приголосних фонем, отриманих з вибірок досліджуваних функціональних стилів. Розглядаються три випадки: 1) позиція фонем в слові є невизначеною; 2) фонема на початку слова; 3) фонема в кінці слова. Дані досліджень для цих трьох випадків наведено у Додатку Е. У таблицях подано значення середніх частот груп приголосних фонем для всіх досліджуваних стилів та відображено два типи ранжування: 1) побудова для кожного стилю спадного ряду значень середніх частот груп приголосних фонем (у цьому випадку найвища середня частота отримує ранговий показник "1", найнижча – "8"); 2) побудова для кожної групи фонем спадного ряду значень середніх частот цієї групи фонем для різних стилів (у цьому випадку найвища середня частота отримує ранговий показник "1", найнижча – "7"). Рангові показники

першого типу ранжування наведено у верхньому лівому кутку кожної клітини. Рангові показник другого типу ранжування подано у правому нижньому кутку кожної клітини (Додаток Е).

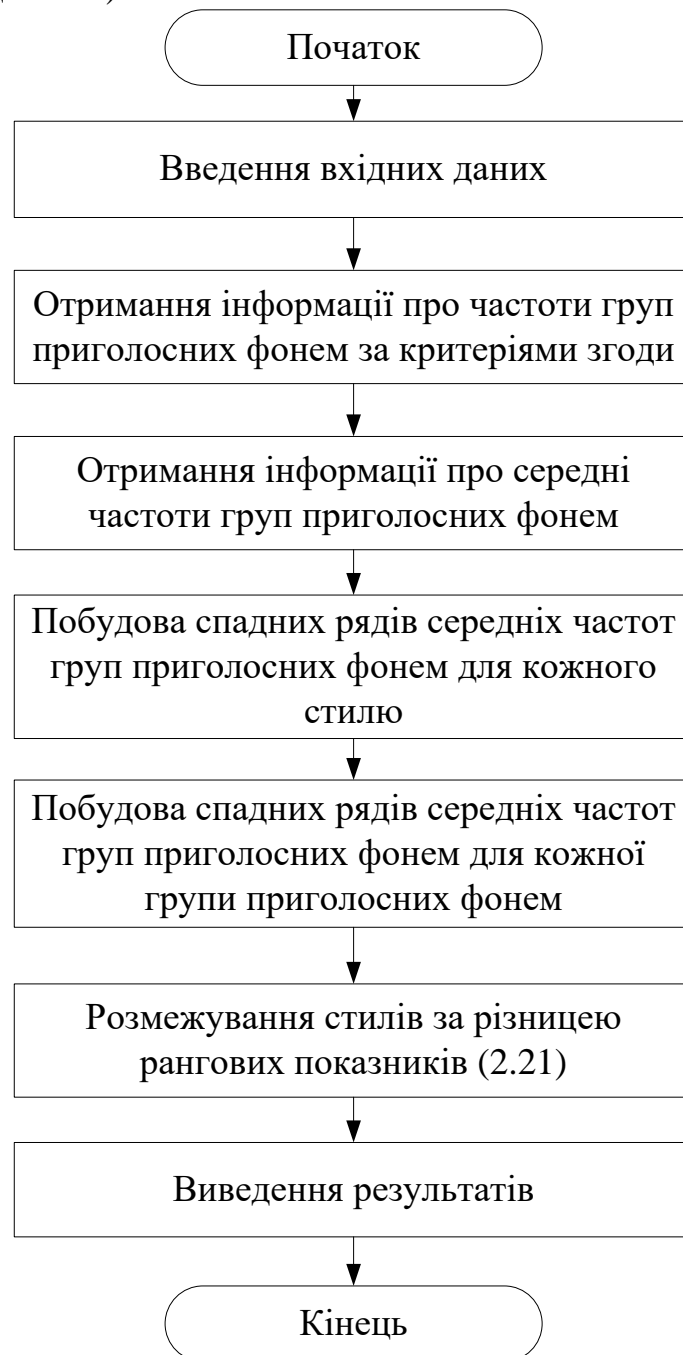


Рисунок 2.4 – Блок-схема алгоритму методу ранжування

У випадку першого типу ранжування одержується розподіл рангових показників величин \bar{X}_r^a по a для кожного стилю, тобто рангові показники вказують номери місць груп приголосних фонем у фонологічній підсистемі системи

досліджуваних стилів. Дані таблиць Додатку Е показують, що розподіл рангових показників груп приголосних фонем однаковий для всіх стилів. У випадку не врахування позиції фонем в слові, ця закономірність виконується майже строго, у випадках позиції фонем на початку та в кінці слова вона виконується менш строго. Цей факт дає змогу стверджувати, що мовний фактор діє сильніше, ніж стилевий. Фонологічний рівень стилю перш за все відображає специфіку фонологічного рівня англійської мови. Стилевий фактор впливає на середні частоти груп приголосних фонем лише у межах, визначених особливостями даної мовної системи, її елементами та зв'язками між ними. Саме зв'язки між елементами системи або її структурний кістяк обмежують дію стилевого фактору, зберігаючи цілісність мовної системи. Відображенням структури мовної системи на числовому рівні є одержані нами середні частоти груп приголосних фонем. Їхня істотна відмінність свідчить про чітко визначене місце у даній мовній системі.

Другий тип ранжування підтверджує правильність вибору середніх частот груп приголосних фонем за критерій диференціації стилів англійської мови. Для кожної групи фонем отримано розподіл рангових показників стилів, який істотно відрізняється від розподілу для іншої групи приголосних фонем. Більша різниця рангових показників стилів вказує на більшу відмінність цих стилів за даною групою приголосних фонем. Для прикладу розглянемо два протилежні випадки:

- 1) максимальна різниця рангових показників;
- 2) мінімальна різниця рангових показників.

1) За групою сонорних для випадку позиції фонем на початку слова зіставлено драму Шоу з науковим стилем. За другим типом ранжування драмі відповідає індекс "1", а науковому стилю – "7". В результаті обчислень, виконаних за методом гіпотез, для даного випадку одержуємо таке значення статистики Стьюдента: $t = 20,57$. Цій величині відповідає дуже мала імовірність: $2Q < 0,1\%$. Звідси отримуємо такий результат: за групою сонорних для випадку позиції фонем на початку слова драма і науковий стиль відрізняються один від одного істотно.

2) За групою сонорних фонем для випадку позиції фонем на початку слова зіставлено драму Шоу з розмовним стилем. За другим типом ранжування драмі відповідає індекс "1", а розмовному стилю – "2". Для цього випадку $t = 1,00: 2Q > 20\%$. За групою сонорних для випадку позиції фонем на початку слова драма і розмовний стиль відрізняються неістотно.

Запропоновано формулу визначення істотних відмінностей між стилями за різницею рангових показників:

$$r_{x_1 - x_2}^{-a} = r_{\max x_1}^{-a} - r_{\min x_2}^{-a}, \quad (2.21)$$

де r – ранговий показник у спадному ряді значень середніх частот груп приголосних фонем; a – група фонем, за якою розмежовуються стилі за різницею максимального та мінімального рангових показників.

Проаналізовано розподіл рангових показників за другим типом ранжування для кожної групи приголосних фонем і наведено найбільш показові результати. Встановлено близькість поезії Дж. Г. Байрона і Т. Мура на фонологічному рівні. У випадку не врахування позиції фонем в слові істотні відмінності цих поезій виявлено за такими групами фонем: сонорні, щілинні, зімкнені. Рангові показники цих груп фонем для поезії Байрона і поезії Мура відрізняються на дві-три одиниці. Рангові показники інших груп фонем відрізняються не більше, ніж на три одиниці. У випадку позиції фонем на початку слова, істотні відмінності виявлено за групами задньоязикових та зімкнених. Рангові показники цих груп фонем для зіставленої поезії відрізняються відповідно на дві і три одиниці. Рангові показники для інших груп фонем відрізняються на одиницю (крім губних – дві одиниці). У випадку позиції фонем в кінці слова істотні відмінності виявлено за групами губних та щілинних фонем. У цьому випадку рангові показники групи губних відрізняються на три одиниці, решти груп фонем – на одиницю.

Отже, близькість поезії Байрона і Мура встановлена не лише за виявленням істотних розходжень за невеликою кількістю груп фонем – дві-три (метод гіпотез), але й за невеликою різницею рангових показників (дві-три одиниці).

Порівняння поезії Байрона і Мура з розмовним стилем також підтверджує їх близькість. У випадку не врахування позиції фонем в слові для поезії Байрона і Мура виявлено істотні розходження з розмовним стилем за тими самими групами фонем: за передньоязиковими, середньоязиковими, задньоязиковими, носовими, щілинними і зімкненими. Для інших груп приголосних фонем різниця рангових показників при зіставленні з розмовним стилем майже однакова: для передньоязикових та середньоязикових – чотири одиниці для поезії Байрона, п'ять одиниць – для поезії Мура, для носових – чотири одиниці для поезії Байрона, три одиниці для поезії Мура; для щілинних – чотири одиниці для поезії Байрона, шість одиниць для поезії Мура. У випадках позиції фонем на початку та в кінці слова, при порівнянні поезії Байрона і Мура з розмовним стилем, істотні розходження встановлено практично за тими самими групами фонем. Різниця рангових показників для окремих груп фонем у цьому порівнянні також майже однакова. Таким чином, застосування методу ранжування при порівнянні поезії Байрона і Мура з розмовним стилем підтверджує слабку відмінність поезії на фонологічному рівні.

Для представлення сильної відмінності стилів наведено результати зіставлень драми Шоу з науковим стилем. Сильна відмінність зіставлених вибірок зумовлена різним складом лексичних одиниць. Так, з одного боку, це емоційно-експресивні елементи та елементи ірландської говірки драми Шоу, а з іншого – наукові терміни з галузі технічної фізики.

Аналіз результатів одержаних за методами гіпотез і ранжування дає змогу ствердити, що у випадку не врахування позиції фонем в слові істотні розходження встановлено за всіма групами фонем, крім губних та носових. При застосуванні методу ранжування одержано такі різниці рангових показників: передньоязикові – 5, середньоязикові – 2, задньоязикові – 4, сонорні – 5, щілинні – 2, зімкнені – 3. У випадку позиції фонем на початку слова істотні розходження встановлено за всіма групами фонем, крім щілинних. Різниця рангових показників: губні – 1, передньоязикові – 3, середньоязикові – 5, задньоязикові – 5, носові – 6, сонорні – 6, зімкнені – 4. У випадку позиції фонем в кінці слова істотні розходження

встановлено за всіма групами фонем, крім носових. Різниця рангових показників: губні – 4, передньоязикові – 4, задньоязикові – 4, сонорні – 4, щілинні – 2, зімкнені – 4. Таким чином, при порівнянні драми Шоу з науковим стилем для всіх трьох випадків позиції фонем в слові істотні розходження встановлено за методом гіпотез і ранжування.

Результати проведеного аналізу підтвердили, що метод ранжування дає істотну інформацію для диференціації стилів англійської мови на фонологічному рівні. За першим типом ранжування визначено дію мовного фактору. Цей тип ранжування показує істотні відмінності середніх частот груп приголосних фонем у межах кожного стилю та відносну стійкість середньої частоти кожної групи приголосних фонем для всіх досліджуваних стилів. Дані результати підтверджують системно-структурну організацію мови, в якій кожна група фонем має своє, чітко визначене місце розташування.

За другим типом ранжування визначено дію стилевого фактору. Побудувавши для кожної групи приголосних фонем спадний ряд середніх частот цієї групи для різних стилів, ми одержали розподіл рангових показників стилів для всіх восьми груп фонем. Велика чи мала відмінність рангових показників зіставлених стилів вказує відповідно на сильну, чи слабку відмінність цих стилів за даною групою приголосних фонем.

Результати одержані за методом ранжування для всіх зіставлених стилів наведено у таблицях Додатку Е (Е.1 – Е.3).

2.4 Використання методу визначення відстаней між стилями для диференціації фоностатистичних структур стилів англійської мови

Відстані між стилями встановлюються у стилеметрії – напрямі досліджень, який спрямований на здійснення вимірів стилевих характеристик. Цей напрям представлено у роботах Г. Н. Мартиненко, С. В. Чебанова [104 – 106].

Для того, щоб встановити відстані між стилями слід виокремити елементарні, доступні виміру ознаки, за якими, незважаючи на складну природу досліджуваного

об'єкта, його можна диференціювати в межах системи однорідних йому об'єктів. Такими елементарними, доступними виміру ознаками у нашому дослідженні є частоти груп приголосних фонем, з яких, внаслідок математичної операції, отримано середні частоти. За середніми частотами груп приголосних фонем визначено відстані між розмовним стилем та іншими функціональними стилями (таблиці Додатку Е (Е.4 – Е.6)).

Подаємо блок-схему алгоритму методу визначення відстаней між стилями (рис. 2.5).

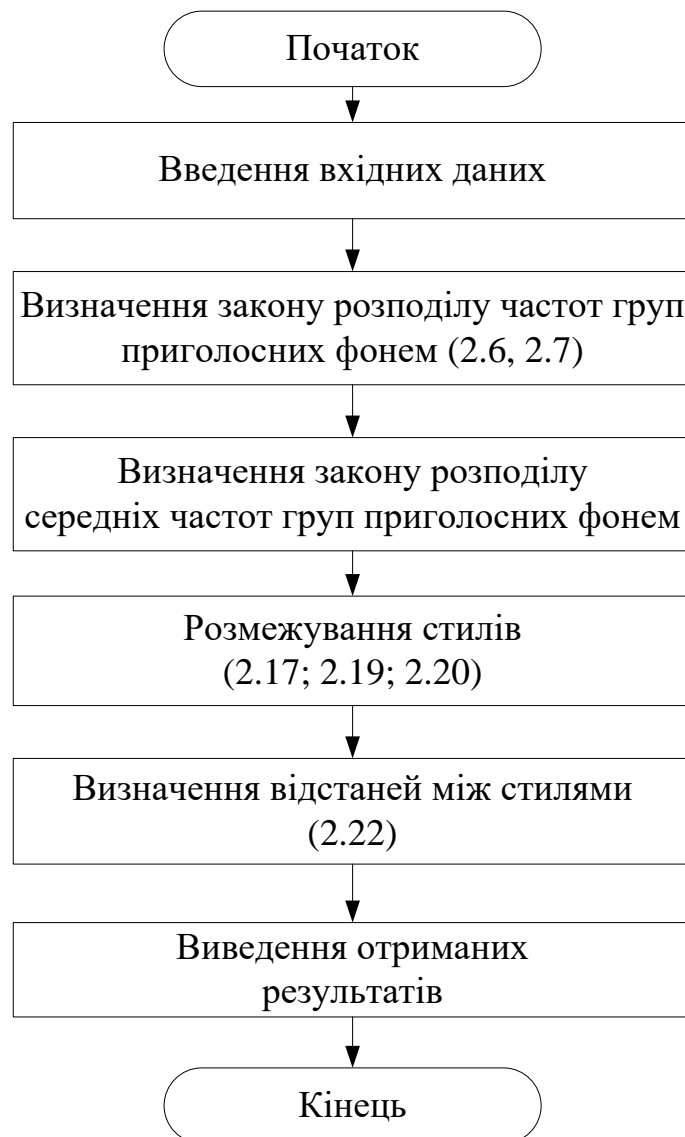


Рисунок 2.5 Блок-схема алгоритму методу визначення відстаней між стилями

Відстані між стилями визначено за методикою, яка передбачає поглиблення та розширення квантитативного підходу у дослідженні мовних явищ [21]:

$$L = \frac{t - t_0}{t}. \quad (2.22)$$

У виразі (2.22) параметр t критерія Стюдента відповідає величині типу $\bar{x}_1 - \bar{x}_0$, t_0 відповідає вибраному рівню значущості, \bar{x}_1 і \bar{x}_0 – середні частоти груп приголосних фонем відповідно певного стилю і розмовного стилю. При визначенні характеристики L використано рівень значущості 0,05.

У дослідженні кількість ступенів вільності $\nu = 60$. Отже, одержано $t_0 = t_{0,05; 60} = 2,00$ [93]. У таблицях Додатку Е (Е.4 – Е.6) наведено відстані досліджуваних стилів від розмовного стилю для випадків не врахування позиції фонем в слові, позиції фонем на початку і в кінці слова. Прочерк означає, що $t < t_0$, і відстані між даним стилем і розмовним стилем немає.

2.5 Розроблення багатofакторного методу для диференціації фоностатистичних структур стилів англійської мови

Розроблений багатofакторний метод передбачає визначення значень ступенів дії трьох факторів (стилю, підстилю та авторської манери викладу), що дає змогу підвищити достовірність стилевої, підстилевої та авторської атрибуції досліджуваного тексту, яка здійснюється за різницею значень загальної стилевої маркованості зіставлюваних текстів.

Розроблено схему стиль→підстиль→автор для встановлення авторства текстів різної тематики, за якою визначається загальна стилева маркованість досліджуваного тексту:

$$sm = \frac{t_{f_1} + t_{f_2} + t_{f_3}}{3}, \quad (2.23)$$

де t – значення критерію Стюдента для різних авторів, стилів і підстилів. Багатofакторний метод визначає ступінь дії факторів авторської манери викладу, стилю та підстилю з довірчою ймовірністю 0,95. Ступінь дії фактора авторської

манери викладу (t_{f_1}) визначається при зіставленні текстів одного стилю, одного підстилю, але різних авторів. Зіставлені тексти повинні належати до одного історичного періоду, літературного напрямку та мати схожу тематику. У цьому випадку нейтралізується дія факторів історичних змін мовлення, характерних особливостей певного літературного напрямку та відмінностей лексико-семантичних груп. У дослідженні зіставлено тексти поезії Байрона і Мура, які належать до одного історичного періоду, літературного напрямку – романтизму та є, до певної міри, тематично схожими. У даному зіставленні текстів визначено схожість за методами гіпотез, ранжування та визначення відстаней між стилями для трьох випадків позиції фонему в слові.

Ступінь дії фактора підстилю (t_{f_2}) визначено у зіставленні текстів підстилів художнього стилю (поезії і художньої прози Байрона, поезії Мура, драми Шоу) за методами гіпотез, ранжування та визначення відстаней між стилями для трьох випадків позиції фонему в слові.

Ступінь дії фактора стилю (t_{f_3}) визначено у зіставленні текстів художнього, розмовного, газетного, публіцистичного та наукового стилів англійської мови за методами гіпотез, ранжування та визначення відстаней між стилями для трьох випадків позиції фонему в слові.

Встановлено ступінь дії факторів авторської манери викладу ($t_{f_1} = \bar{x}_{1a}^a - \bar{x}_{2a}^a$), підстилю ($t_{f_2} = \bar{x}_{1n}^a - \bar{x}_{2n}^a$) та стилю ($t_{f_3} = \bar{x}_{1c}^a - \bar{x}_{2c}^a$). Визначаємо середнє від трьох значень t . Визначене середнє значення є значенням загальної стилєвої маркованості досліджуваного тексту sm (2.23).

Відмінність між авторськими стилями визначається за різницею значень загальної стилєвої маркованості:

$$sm_1 - sm_2,$$

де sm_1 – значення загальної стилєвої маркованості тексту першого автора, а sm_2 – значення загальної стилєвої маркованості тексту другого автора.

Розроблений алгоритм методу представлено на рис. 2.6.

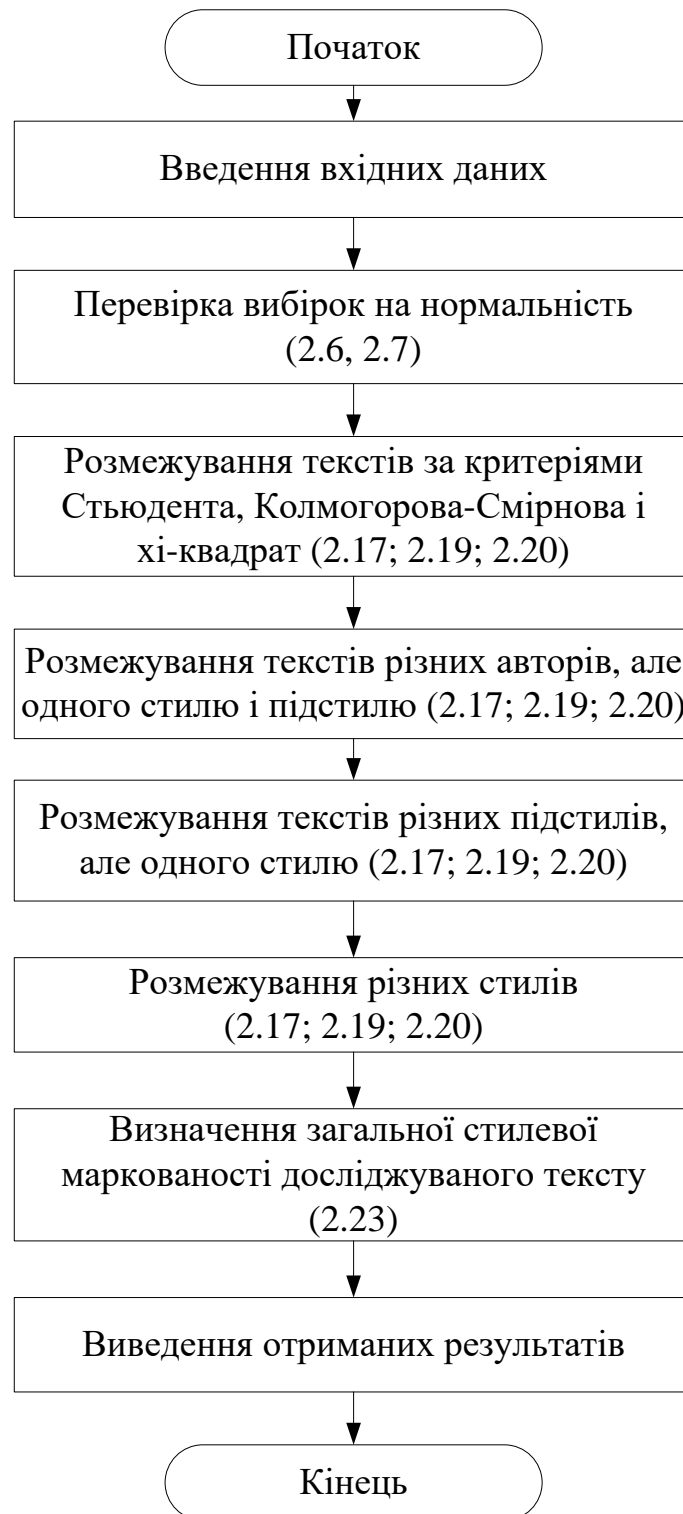


Рисунок 2.6 – Блок-схема алгоритму багатofакторного методу

Розроблений метод дає змогу з більшою достовірністю встановити загальну рівнодіючу факторів стилю, підстилю та авторської манери викладу та здійснити авторську атрибуцію досліджуваного тексту за різницею значень загальної стилеві маркованості зіставляваних текстів.

2.6 Висновки до розділу 2

1. Вперше розроблено метод комплексного аналізу диференціації фоностатистичних структур стилів англійської мови, який ґрунтується на запропонованому поєднанні таких статистичних методів: методу гіпотез, методу ранжування та методу визначення відстаней між стилями. Метод гіпотез (критерій Стюдента) передбачає перевірку відповідності частот груп приголосних фонем нормальному розподілу (критерій Пірсона, спрощений критерій Романовського); встановлення того, що середні частоти груп приголосних фонем підлягають нормальному розподілу (теорема Ляпунова), встановлення обсягу вибірки, який є достатнім для одержання достовірної інформації про стилі, диференціацію стилів (критерій Стюдента, критерій Колмогорова-Смірнова, критерій χ^2 -квадрат). Метод ранжування передбачає встановлення істотних відмінностей між стилями за різницею рангових показників (запропоновано формулу). Метод визначення відстаней між стилями розроблено В. С. Перебийніс. Запропоноване поєднання розглянутих методів дає змогу підвищити достовірність розмежування стилів.

2. Вперше розроблено багатofакторний метод визначення ступенів дії трьох факторів (авторської манери викладу, підстилю та стилю), який уможливорює підвищення достовірності встановлення авторства тексту за загальною стилевою маркованістю досліджуваного тексту з довірчою ймовірністю 0,95.

3. Розроблено алгоритми практичної реалізації методу комплексного аналізу диференціації фоностатистичних структур стилів англійської мови, складовими якого є: метод гіпотез, метод ранжування та метод визначення відстаней між стилями; багатofакторний метод для розмежування текстів різних авторів. Побудовані алгоритми забезпечують швидку програмну реалізацію розроблених методів та автоматизацію процесів авторської атрибуції англомовних текстів.

РОЗДІЛ 3. МОДЕЛЬ СИСТЕМИ ФОНОСТАТИСТИЧНИХ СТРУКТУР ДОСЛІДЖУВАНИХ СТИЛІВ АНГЛІЙСЬКОЇ МОВИ

3.1 Вдосконалення статистичних моделей – складових моделі системи фоностатистичних структур стилів

У дисертаційному дослідженні побудовано модель системи фоностатистичних структур досліджуваних стилів (художнього, розмовного, газетного та наукового) англійської мови, яка є аналогом системи функціональних стилів англійської мови. Складовими цієї моделі системи 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 методом гіпотез і ранжування, яка на відміну від існуючих, враховує позицію фонем у слові, визначає ступінь встановлених істотних відмінностей за відношенням кількості груп приголосних фонем, за якими встановлено істотні відмінності між попарно зіставленими стилями до загальної кількості груп приголосних фонем.

Ідея диференціації 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. У плані статистичного опису поставленого в роботі завдання – розмежування стилів, слід зазначити, що досліджується деяка випадкова величина X . Результат першого спостереження – це випадкова величина, яка позначається X_1 , результат другого спостереження – X_2 , результат n -го спостереження – X_n . Сукупність усіх спостережень – це випадковий вектор (X_1, \dots, X_n) , який позначається

\tilde{X} . Тобто $\tilde{X} = (X_1, \dots, X_n)$. Отже, вибірка – це випадковий вектор. Вважається, що результати спостережень є незалежними і однаково розподіленими.

Для дослідження розглядається числовий вектор, бо статистичними даними є певні числові значення. Ці числові значення позначаються x_1, \dots, x_n , їх сукупність – числовий вектор $\tilde{x} = (x_1, \dots, x_n)$. Вектор \tilde{x} – це реалізація вибірки. Оскільки у роботі зіставляються дві вибірки, то вибірка X – це вектор складений із кількості фонем певної групи на кожному порцію (1000 фонем) з 51 порції у першому тексті, а вибірка Y – це вектор складений із кількості фонем цієї ж групи у другому тексті. Для кожної групи фонем є дві реалізації вибірок з двох текстів.

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

Крок 1. Встановити істотні відмінності для 8 груп приголосних фонем у зіставленні стилів, підстилів і текстів різних авторів, використовуючи метод гіпотез: критерій Стюдента (2.17), Колмогорова-Смірнова (2.19) та критерій хі-квадрат (2.20) з довірчою ймовірністю 0,95.

Крок 2. Розмежувати стилі, підстилі та тексти різних авторів за встановленими істотними відмінностями, використовуючи метод ранжування (2.21).

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):

$$DSD = \frac{SDGN}{TGN}. \quad (3.1)$$

Крок 5. Встановити низький ступінь встановлених істотних відмінностей ($LDSD$) за відношенням 1 – 3 груп приголосних фонем, за якими попарно зіставлені

стилі відрізняються істотно, 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. Встановити середній ступінь встановлених істотних відмінностей (*MDSN*) за відношенням 4 – 5 груп приголосних фонем, за якими встановлено істотні відмінності між попарно зіставленими стилями, 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 восьми 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 розмежування стилів за критеріями перевірки гіпотези на однорідність вибірок та за різницею значень рангових показників. Якщо кількість груп приголосних фонем, за якими встановлено істотні відмінності між попарно зіставленими стилями дорівнює загальній кількості груп приголосних фонем, то ступінь розмежування стилів є найвищим. Якщо кількість груп приголосних фонем, за якими встановлено істотні відмінності між попарно зіставленими стилями дорівнює 1, то ступінь розмежування 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 за певною групою приголосних фонем, до кількості всіх порівнянь за всіма групами приголосних фонем.

Зменшення кількості груп приголосних фонем, за якими диференціюються стилі, спрощує процес атрибуції тексту і підвищує його достовірність 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 Стьюдента (2.17), Колмогорова-Смірнова (2.19) та критерієм хі-квадрат (2.20).

Крок 2. Визначити групу фонем, за якою встановлено істотні відмінності за критерієм Стьюдента, Колмогорова-Смірнова та критерієм хі-квадрат:

$$ADC_1 = \frac{t + \lambda_{n,m} + \hat{\chi}_n^2}{3}, \quad (3.2)$$

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

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

$$ADC_2 = \frac{UP + IP + FP}{3}, \quad (3.3)$$

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):

$$ADC = \frac{SSD}{TNC}. \quad (3.4)$$

Крок 5. Визначити групу фонем з найбільшим значенням *ADC*.

Крок 6. Розмежувати тексти за цією групою фонем.

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

У дисертаційному дослідженні вдосконалені статистичні моделі застосовано для розмежування текстів, які належать до різних функціональних стилів. Відповідно до загальноприйнятої термінології досліджувані тексти є вибірками, які є частинами генеральних вибірок – сукупностей всіх існуючих текстів кожного із досліджуваних функціональних стилів.

Усі досліджувані вибірки мають однакову величину. Однакова величина вибірки є умовою отримання достовірних результатів, бо в такий спосіб можна уникнути впливу на частоту фонем додаткового фактора, пов'язаного з різною величиною вибірки.

Важливим аспектом дослідження є підбір тексту за стилевими характеристиками та тематикою. Вдалий підбір тексту допомагає встановити статистичні характеристики окремих фонем та їх груп в межах певної фонологічної системи. Це, у свою чергу, дасть змогу встановити, чи частота фонем залежить від стилістичних характеристик даної мовної системи.

Необхідність виокремлення в певній мовній системі певної кількості функціональних стилів зумовлює їх виокремлення на всіх рівнях мовної системи, а отже, і на фонологічному.

Розроблена блок-схема алгоритму зіставлення досліджуваних стилів із розмовним стилем зображена на рис. 3.1).

Модель системи фоностатистичних структур стилів визначає співвідношення художнього, газетного і наукового стилів із фоностатистичною структурою розмовного стилю.

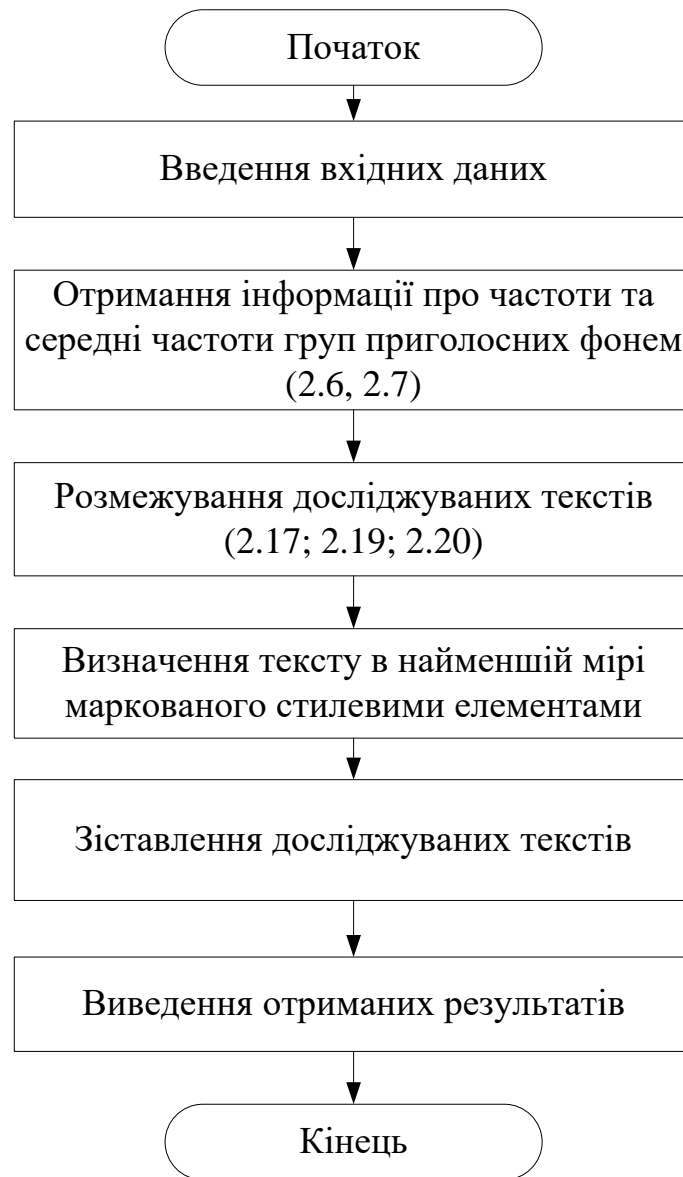


Рисунок 3.1 – Блок-схема алгоритму зіставлення досліджуваних стилів із розмовним стилем

Вибір текстів для зіставлення повинен враховувати їх специфіку. Чи слід порівнювати тексти, що належать до різних функціональних стилів попарно один з одним, чи доцільніше порівнювати їх з умовно вибраним "нейтральним" текстом? Про необхідність порівняння з "нейтральним" текстом висловлювався Н. С. Трубецької [107]. Неможливість чітко визначити "нейтральний" стиль зумовлена взаємопроникненням елементів функціональних стилів. Крім того, абсолютно нейтрального стилю взагалі не існує, бо кожен із функціональних стилів

є маркованим, в тій чи іншій мірі, емоційно-експресивними елементами, чи елементами, характерними для окремих сфер людської діяльності. Проте доречно виокремити стиль, який в найменшій мірі є маркованим емоційно-експресивними елементами (на відміну від художнього стилю) та спеціальною термінологією (на відміну від наукового та газетного стилів). Таким стилем можна умовно вважати літературний різновид розмовного стилю.

Слід зауважити, що взаємопроникнення стилів ускладнює чітке розмежування функціональних стилів. Цим пояснюється виокремлення різної кількості стилів різними дослідниками та різнобій у визначенні приналежності певного тексту до певного функціонального стилю. Така неоднозначність у трактуванні стилю залежить і від суб'єктивно вибраних дослідником критеріїв, і від об'єктивних причин – насамперед складної природи мови, яка є відкритою системою, в яку проникають нові елементи (неологізми), і деякі елементи виходять із використання (архаїзми).

Питання величини вибірки є питанням великої ваги для дослідників, зацікавлених в отриманні достовірних статистичних даних, які можуть стати основою для лінгвістичних теоретичних припущень, гіпотез. Тому в багатьох роботах можна натрапити на критичні зауваження щодо недостатньої величини вибірки.

Так, наприклад, підкреслюючи всі позитивні сторони робіт Ванга і Крофорда (про які згадано раніше), не можна погодитися із висновком про можливість використання малих вибірок для статистично-фонологічних обчислень. Торкаючись цієї проблеми, Д. М. Сегал слушно зауважує: "В тих випадках, коли статистика ґрунтується на одиничній, (при цьому малій), вибірці, її ледве, (якщо взагалі), можна вважати достатньою для достовірних результатів" [20]. Ципф у критичному аналізі досліджень Бурдона вважає отримані ним дані фактично недійсними, бо Бурдон досліджував вибірки надто малого обсягу – не більше 3000 елементів [30].

Вибірки великого обсягу використано у дослідженнях Д. М. Сегала, Р. Г. Піотровського та інших дослідників [19, 20]. У дослідженні величина кожної із досліджуваних вибірок становить 51 тисячу фонем. При цьому обсязі вибірки,

інтервал довіри, в якому знаходиться частота генеральної сукупності, є достатньо малий, що дає змогу говорити про конкретне значення середньої частоти групи фонем генеральної сукупності. Тобто отримується достовірна інформація про параметр, за яким розмежовуємо функціональні стилі англійської мови.

3.2 Розроблення алгоритму врахування позиції фонем в слові для моделі системи фоностатистичних структур стилів

Модель системи фоностатистичних структур художнього, газетного, наукового та розмовного стилів розроблено із врахуванням позиції фонем в слові для підвищення достовірності визначення співвідношення фоностатистичних структур стилів.

Частота вживання певної фонем або групи фонем в різних позиціях слова є різною. Істотні відмінності між стилями можна виявити у всіх позиціях фонем в слові, в одній із можливих, в декількох із можливих або в жодній. Тому для одержання достовірних даних про істотні відмінності між зіставленими стилями, необхідно враховувати позицію фонем в слові.

На рис. 3.2 зображено блок-схему алгоритму врахування позиції фонем в слові.

У дослідженні враховується позиція фонем у слові. Крім середньої частоти у випадку не врахування позиції фонем в слові, визначено середні частоти для позиції фонем на початку та в кінці слова. Розглянуті три випадки дають змогу зробити висновок про залежність стилерозрізняльної здатності фонем від її позиції в слові. Перш за все, ця залежність пояснюється фонологічною природою кожної фонем – складом диференційних ознак, які визначають частоту вживання фонем в певних позиціях в слові. Тому для певних фонем, частота вживання вища на початку слова, для інших, навпаки, нижча. Кожна фонема характеризується певною частотою вживання для всіх трьох випадків. Наприклад, фонема *h* не вживається в кінці слова, і тому для цієї позиції стилерозрізняльна здатність фонем *h* є нульовою. Отже, для того, щоб визначити ті розходження, які є зумовленими

стилевими особливостями текстів, необхідно враховувати позиції, в яких фонема повністю розкриває свою специфіку частоти вживання.

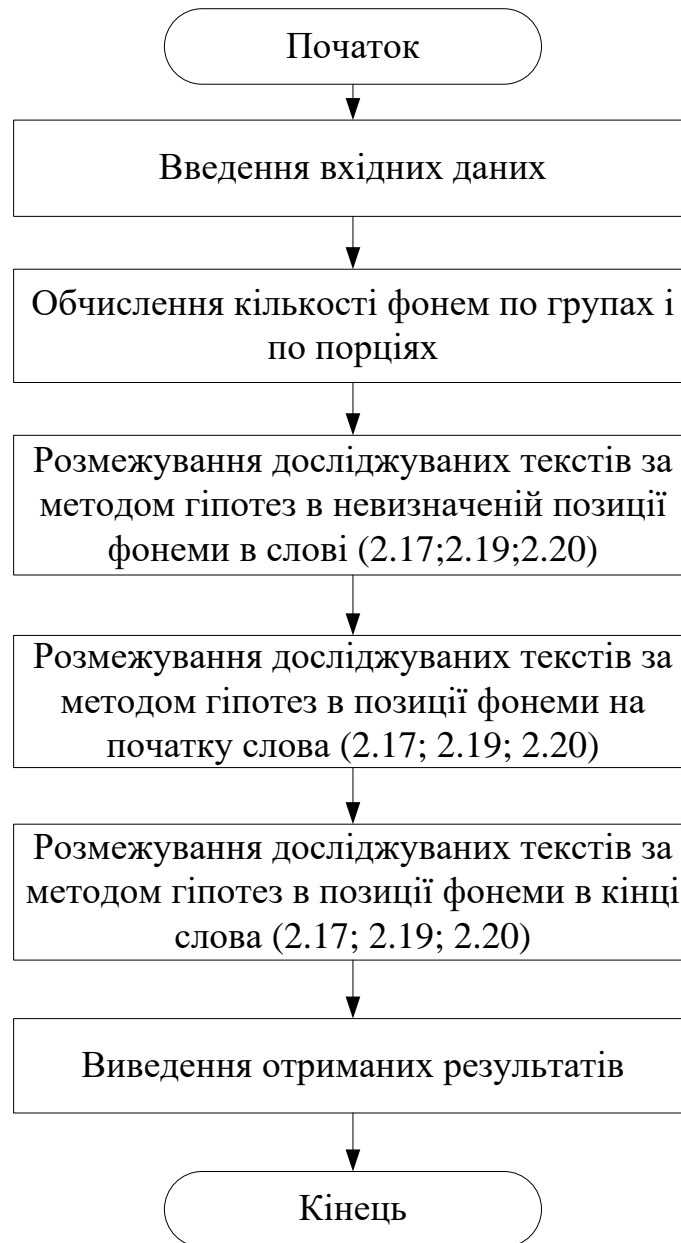


Рисунок 3.2 – Блок-схема алгоритму врахування позиції фонем в слові

3.3 Розроблення алгоритму визначення частотних характеристик груп приголосних для моделі системи фоностатистичних структур стилів

Модель системи фоностатистичних структур художнього, газетного, розмовного та наукового стилів розроблено на основі визначення частотних характеристик груп приголосних фонем. Кожна група приголосних фонем має свої

частотні характеристики, які визначаються у попарному зіставленні стилів для різних випадків позицій фонемі в слові. Визначення частотних характеристик груп приголосних фонем дає інформацію про істотні відмінності між стилями.

Статистичне обстеження текстів, які представляють певні функціональні стилі, можна проводити у двох основних напрямках. Перший напрям передбачає визначення частоти кожної фонемі, за якою можна визначити місце кожної фонемі в певній фонологічній системі. В основі другого напрямку є порівняння частот фонемі, встановлених для різних функціональних стилів, а також визначення відмінностей між ними. Отже, у першому випадку отримується розподіл частот фонем, характерний для певної фонологічної системи, у другому визначається, наскільки істотно стилеві особливості текстів впливають на частоту фонем.

Слід зазначити, що абстрагування від стилевих особливостей при визначенні суто мовних характеристик є умовним. Бо очевидним є те, що абсолютно позбавленого стилевих особливостей, тобто "нейтрального" тексту, не існує. Таким чином, одержані показники відображають і мовні, і стилеві особливості кожної окремої фонемі одночасно.

Дослідження у двох вищезгаданих напрямках можна проводити як для кожної фонемі зокрема, так і для груп фонем. У дослідженні статистичними параметрами функціональних стилів є середні частоти груп приголосних фонем. Отже, дослідження проведено у другому напрямі.

Приголосні фонемі об'єднуються у групи за певними диференційними ознаками, які відбивають артикуляційні особливості фонем. Відповідно до класифікації за місцем творення, виокремлюють:

- губні: p, b, m, f, v, w;
- передньоязикові: θ, ð, t, d, n, s, z, l, tʃ, dʒ, ʃ, ʒ, r;
- середньоязикові: j;
- задньоязикові: k, g, ŋ.

За способом творення приголосні поділяють на: носові: m, n, ŋ; сонорні: w, r, j, l; щілинні: f, v, θ, ð, h, s, z; зімкнені: p, b, t, d, k, g, tʃ, dʒ [108].

Групування за певними диференційними ознаками свідчить про наявність певних відношень, зв'язків між фонемами, тобто певної структури, яка є характеристикою системи фонем. Фонем об'єднуються в групи за ознаками, які є спільними для даної групи, і за якими вони водночас протиставляються всім іншим фонемам, що належать до інших груп.

Блок-схема алгоритму визначення частотних характеристик груп приголосних фонем зображена на рис. 3.3.

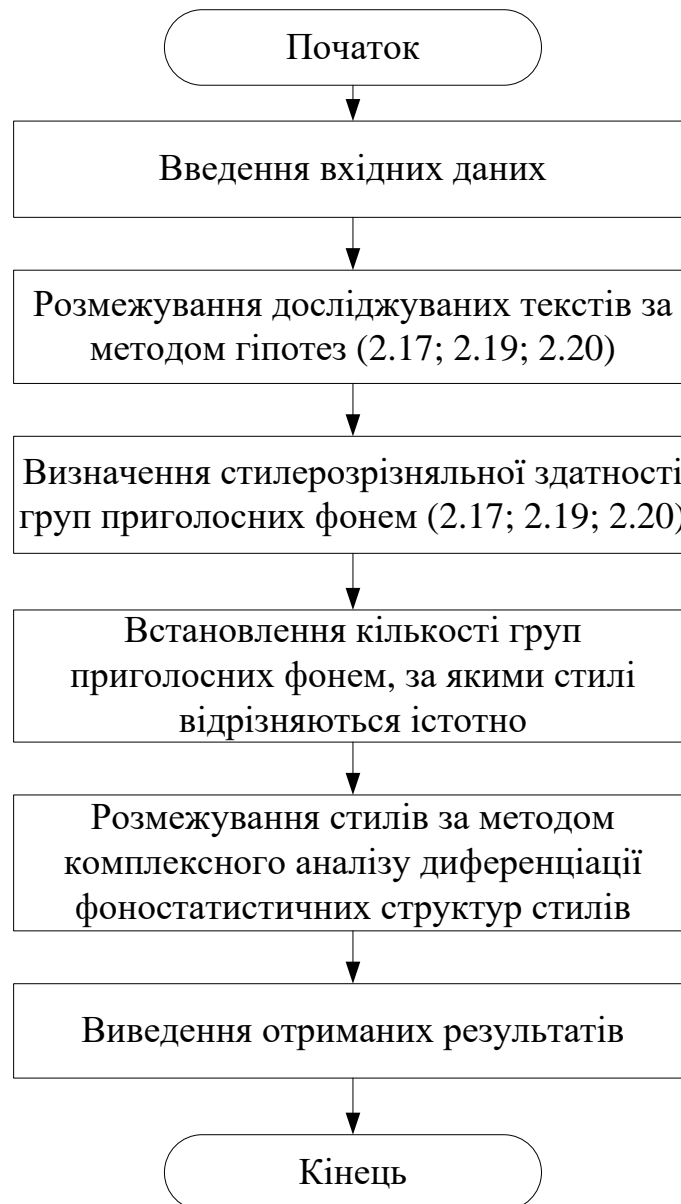


Рисунок 3.3 – Блок-схема алгоритму визначення частотних характеристик груп приголосних фонем

Отримані числові дані для вищезазначених груп приголосних фонем з математичною точністю підтверджують системно-структурний характер мови, бо вони кількісно показують відмінності між групами приголосних фонем, встановлюють місце кожної групи приголосних фонем у фонологічній системі англійської мови.

Експериментально встановлена різна здатність груп приголосних фонем розмежовувати стилі свідчить про відмінність функціональної ролі кожної з цих груп у межах певної фонологічної системи.

Варто зауважити, що для встановлення розходжень між стилями на фонологічному рівні, зовсім не обов'язково виявити істотні розходження за всіма групами приголосних фонем. Припустимо, що для деяких з них, розходження є неістотними. Встановлення неістотних розходжень не означає встановлення відсутності розходжень взагалі. Незначні розходження є частотними характеристиками груп приголосних фонем.

У дисертаційному дослідженні отримано істотні розходження для більшості груп приголосних фонем, і тим самим, визначено статистичні параметри стилів. Саме ці статистичні параметри, або набір статистичних характеристик для кожного із стилів, є об'єктивним критерієм для стилістів, за допомогою якого можна з більшою точністю охарактеризувати кожен функціональний стиль. Статистичним параметром, за яким розмежовано функціональні стилі за критерієм Стюдента, є середня частота груп приголосних фонем, яка є відносно стійкою. Відносна стійкість середньої частоти груп приголосних фонем для певного функціонального стилю є умовою його відмежування від інших функціональних стилів у межах певної фонологічної системи. Свідомий добір мовних одиниць, за яким формуються тексти різних функціональних стилів зумовлює стилеві характеристики текстів, виражені коливанням частоти фонем в межах середньої частоти. Зрозуміло, що свідомий добір торкається фонологічного рівня в найменшій мірі, оскільки, у більшості випадків, автор добору орієнтується на смислове, а не звукове наповнення слів. Фонемі, як складові елементи звукової оболонки слова, не враховуються в ході добору, хіба за винятком тих випадків, коли автор орієнтується на звучання слова з

метою досягнення певного звукового експресивного ефекту. Наприклад, у деяких зразках поезії. Отже, говорити про абсолютну відсутність добору на фонологічному рівні теж недоцільно. Такий добір, певною мірою, впливає на частоту вживання фонем.

3.4 Побудова алгоритму опису рівнів мови та мовлення для моделі системи фоностатистичних структур стилів

Модель системи фоностатистичних структур художнього, газетного, наукового та розмовного стилів розроблено із врахуванням рівнів мови і мовлення. Визначення середньої частоти груп фонем дало змогу чітко розмежувати рівні мови і мовлення.

Рівень мовлення представляють частоти окремих порцій груп приголосних фонем. Рівень мови представляють середні частоти груп приголосних фонем, які отримано в ході математичної операції визначення середньої частоти.

Отже, чітко відстежується взаємозалежність двох рівнів, бо середні частоти груп приголосних фонем одержуються шляхом абстрагування від численних частот окремих порцій груп приголосних фонем і тому безпосередньо залежать від них. Інакше кажучи, спостерігається залежність рівня мови від рівня мовлення.

Зворотна залежність рівнів виявляється у тому, що численні частоти окремих порцій певної групи фонем коливаються у межах середньої частоти цієї групи, що, в свою чергу, підтверджує наявність структурних зв'язків у мовній системі.

Блок-схема алгоритму опису рівнів мови та мовлення у статистичній системі фоностатистичних структур стилів зображена на рис. 3.4.

Таблиця 3.1 показує співвідношення числових даних на рівнях мови і мовлення для групи губних у вибірці поезії Дж. Г. Байрона. Числові дані, які представляють рівень мовлення, є частотами окремих порцій групи губних фонем для кожної із 31 порції вибірки. Числове значення, яке представляє рівень мови, є середньою частотою групи губних фонем для даної вибірки.

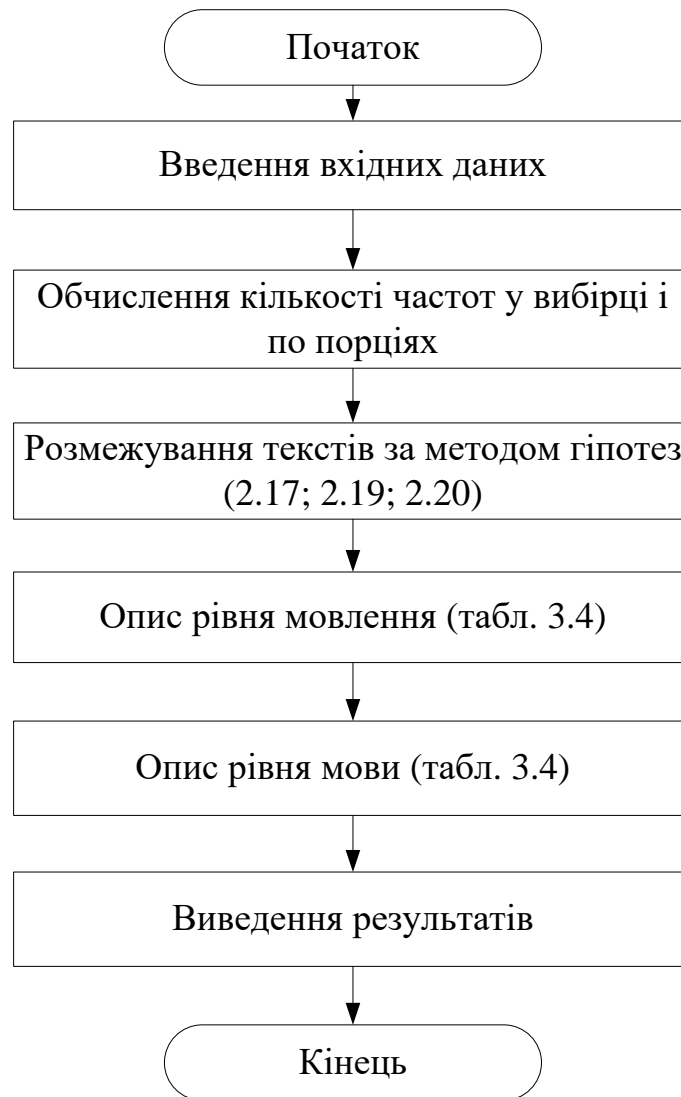


Рисунок 3.4 – Блок-схема алгоритму опису рівнів мови та мовлення у моделі системи фоностатистичних структур стилів

Порівнюючи фонологічну і статистичну системи (систему фоностатистичних структур стилів), співставляються не лише елементи цих систем, але й відношення між цими елементами. Чи існують аналогі парадигматичних (внутрішньокласових) та синтагматичних (міжкласових) відношень у статистичній системі? Для того, щоб позитивно відповісти на це запитання, необхідно визначити середню частоту вживання кожної фонемі. Одержання середнього значення частоти певної фонемі здійснюється математичною операцією, у якій ми абстрагуємось від конкретних значень частоти вживання певної фонемі. Відхилення у певних допустимих межах від визначеної середньої частоти, зумовлено впливом внутрішніх – мовних і

зовнішніх – стилевих чинників. Навколо середньої частоти, частоти фонем можна об'єднати у класи або парадигми. У межах парадигми кожної фонемі парадигматичні відношення є відношеннями між середньою частотою фонемі і частотами численних реалізацій фонемі в мовленні.

Таблиця 3.1

Співвідношення числових даних на рівнях мови і мовлення для групи губних у вибірці поезії Дж. Г. Байрона

Рівень мовлення						Рівень мови
№ порції	Частота	№ порції	Частота	№ порції	Частота	Середня частота
1	135	11	158	21	141	134
2	128	12	120	22	128	
3	133	13	130	23	124	
4	140	14	137	24	135	
5	123	15	134	25	113	
6	150	16	127	26	141	
7	146	17	137	27	145	
8	133	18	131	28	133	
9	148	19	134	29	125	
10	134	20	137	30	121	
				31	142	

Парадигми окремих фонем об'єднуються у більші парадигми за певними артикуляційно-акустичними ознаками. Парадигми такого порядку є об'єктом даного дослідження. Це групи губних, передньоязикових, середньоязикових, задньоязикових, носових, сонорних, зімкнених та щілинних фонем. Процес розширення парадигм веде до утворення парадигм ще вищого порядку – парадигм приголосних та голосних і завершується утворенням надпарадигми – парадигми фонем англійської мовної системи.

Відношення між частотами у межах кожної парадигми є парадигматичними відношеннями, а відношення між парадигмами є ієрархічними відношеннями.

Розглядаючи специфіку синтагматичних відношень у моделі статистичної системи, слід зазначити, що це відношення між частотами фонем, які межують одна з одною у мовленнєвому ланцюгу. Слід зауважити, що на рівні мовлення виявляються як синтагматичні, так і парадигматичні відношення, проте перші

виявляються реально, останні – потенційно. Потенційним виявленням парадигматичних відношень вважається коливання частот кожної фонемі у межах середньої частоти, яка є класоутворювальною або парадигмоутворювальною. Така кореляція між фонологічною та статистичною системами свідчить про те, що методи математичної статистики дають змогу одержати істотну інформацію про специфіку функціонування фонологічної системи.

Вищерозглянуту модель статистичної системи можна вважати своєрідною моделлю фонологічної системи, в якій у числовому вираженні відображено специфіку фонологічної системи. У дослідженні надпарадигмою є не фонологічна система англійської мови, а фонологічна система приголосних фонем англійської мови.

Розглянемо парадигматичні відношення статистичної системи. Це відношення між середніми частотами груп приголосних фонем, виокремлених за артикуляційно-акустичними ознаками. Середні частоти груп приголосних фонем порівнюються попарно між собою за принципом приналежності до різних функціональних стилів. Так, наприклад, середня частота групи губних фонем, отримана при дослідженні текстів художнього стилю, порівнюється із середніми частотами відповідних груп губних фонем текстів розмовного, газетного та наукового стилів. При порівнянні кожного із стилів із розмовним стилем та попарно між собою за середніми частотами восьми груп приголосних фонем, отримано розподіл середніх частот, який підтверджує, або заперечує близькість чи віддаленість тих чи інших стилів. Отже, статистичні параметри функціональних стилів (наведені у підрозділах 4.2 – 4.5) встановлено завдяки відстеженню парадигматичних відношень. Ці парадигматичні відношення складають статистичну структуру текстів, важливість якої неодноразово підкреслювалась дослідниками. Зокрема Л. Р. Зіндер зазначає: "... лише статистична структура тексту, як деяка рівнодіюча, дає змогу встановити приналежність тексту до відповідного функціонального стилю..."[109].

Об'єктивним фактом є первинність фонологічної системи щодо статистичної. Обидві системи отримано шляхом абстрагування і до певної міри є ідеальними об'єктами, однак статистична система є похідною стосовно фонологічної системи.

Різна частота фонем зумовлена різним складом диференційних ознак, тобто саме мовним чинником. Використання моделі статистичної системи дослідником дає змогу співставити її з фонологічною системою, встановити взаємовідношення між якісними та кількісними характеристиками фонем, висловити теоретичні припущення та пояснити ті чи інші явища мови.

Наявність статистичних даних, отриманих за допомогою моделі системи фоностатистичних структур стилів для кожного із функціональних стилів, допомагає стилістові точніше встановити характерні особливості стилів у межах даної мовної системи. Взаємовідношення між якісними та кількісними характеристиками, звичайно, визначаються пріоритетом перших. Справді, маючи лише дані, отримані за допомогою моделі системи фоностатистичних структур стилів, тобто певні числові значення, не можна відтворити конкретну фонологічну систему. Проте не слід вважати, що числові дані можуть лише експериментально підтвердити якісні характеристики функціонування фонем в межах даної фонологічної системи, чи в різних функціональних стилях. Числові величини, в деяких випадках, допомагають дослідникові пояснити ті якісні характеристики, які важко виявити без застосування статистики.

Отже, одержані на синтагматичному рівні частоти, через математичне опрацювання, тобто визначення середньої частоти груп фонем, стають показниками розподілу функціональних стилів на фонологічному рівні.

3.5 Висновки до розділу 3

1. Розроблено модель системи фоностатистичних структур стилів англійської мови (художнього, розмовного, газетного публіцистичного та наукового), яка є вторинною стосовно фонологічної системи. Модель є аналогом системи функціональних стилів англійської мови. Складовими цієї моделі 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. Рівень мови моделі системи

фоностатистичних структур стилів представляють середні частоти груп приголосних фонем, а рівень мовлення – частоти окремих порцій груп приголосних фонем. Відстеження парадигматичних відношень (рівень мови) між середніми частотами груп приголосних фонем, одержаних з вибірок різних функціональних стилів, дало змогу встановити статистичні параметри стилів. Статистичним параметром, за яким можна диференціювати стилі, є середня частота груп приголосних фонем – відносно стійка величина і мотивована стилевими особливостями текстів. Визначення середньої частоти груп приголосних фонем як статистичного параметру розрізнення стилів є спробою розв'язати завдання визначення границь функціональних стилів з більшою достовірністю. Відносна стійкість середньої частоти груп приголосних фонем для певного функціонального стилю є умовою його відмежування від інших функціональних стилів у межах певної фонологічної системи. Функціональні стилі досліджено на матеріалі 14 вибірок обсягом в 51 тисячу фонем. Цей обсяг є достатньо великим для отримання достовірних результатів. Повну характеристику стилерозрізняльної здатності груп приголосних фонем визначено на основі числових даних одержаних для наступних випадків: не врахування позиції фонем в слові, позиція фонем на початку слова, позиція фонем в кінці слова. Встановлене взаємовідношення між фонологічною і статистичною системами розкриває специфіку закономірностей функціонування фонологічної системи і дає змогу зробити теоретичні припущення щодо приналежності досліджуваного тексту до певного стилю, підстилю та встановити його авторство.

2. Розроблено алгоритми реалізації: статистичної моделі стилевої, підстилевої та авторської диференціації текстів за методами гіпотез і ранжування; статистичної моделі визначення стилерозрізняльної здатності груп приголосних фонем досліджуваних стилів.

3. Побудовано алгоритм зіставлення досліджуваних стилів із розмовним стилем, алгоритм врахування позиції фонем в слові, алгоритм визначення частотних характеристик груп приголосних фонем та алгоритм опису рівнів мови та мовлення в моделі системи фоностатистичних структур стилів англійської мови.

РОЗДІЛ 4. РОЗРОБЛЕННЯ ПРОГРАМНОЇ СИСТЕМИ ДИФЕРЕНЦІАЦІЇ ФНОСТАТИСТИЧНИХ СТРУКТУР ФУНКЦІОНАЛЬНИХ СТИЛІВ АНГЛІЙСЬКОЇ МОВИ ТА РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

4.1 Розроблення структури системи диференціації фоностатистичних структур функціональних стилів англійської мови

Розроблено програмну систему автоматизації диференціації фоностатистичних структур функціональних стилів англійської мови.

Програмну систему розроблено з використанням об'єктно-орієнтованої мови програмування Java. Програму написано в середовищі програмування IntelliJIdeaCommunity 14 від розробника JetBrains.

На відміну від комерційної версії IntelliJIDEAUltimate Community має обмежену функціональність, але достатньо функцій для створення цієї програми. Дане середовище розробки має широкий набір інтегрованих інструментів, що дає змогу програмісту швидко реорганізувати ядро програми.

Для реалізації графічного інтерфейсу користувача GUI використано бібліотеку Java під назвою Swing. Swing побудовано з використанням AWT (AbstractWindow Toolkit) [110].

Для розроблення програми використано бібліотеку вводу і виводу JavaI/O. Бібліотека заснована на потоках. Потоки зручно використовувати для читання і запису даних з файлу, мережі, консолі та іншого джерела. Відфільтровані потоки символів дають змогу виконати стиснення, переклад та інші операції.

Для опису математичних формул з теорії ймовірностей використано спеціальну бібліотеку Colt. Colt розроблено для розв'язання завдань фізики високих енергій, лінійної алгебри, статистики та інших завдань. Colt забезпечує інфраструктуру для науково-технічних обчислень у Java [111].

Розроблено структуру програмної системи диференціації фоностатистичних структур функціональних стилів англійської мови (рис. 4.1), яка ґрунтується на модульному принципі, що дає змогу швидко модифікувати розроблене програмне

забезпечення, підвищити рівень автоматизації та ефективність стилевої, підстилевої і авторської атрибуції тексту.

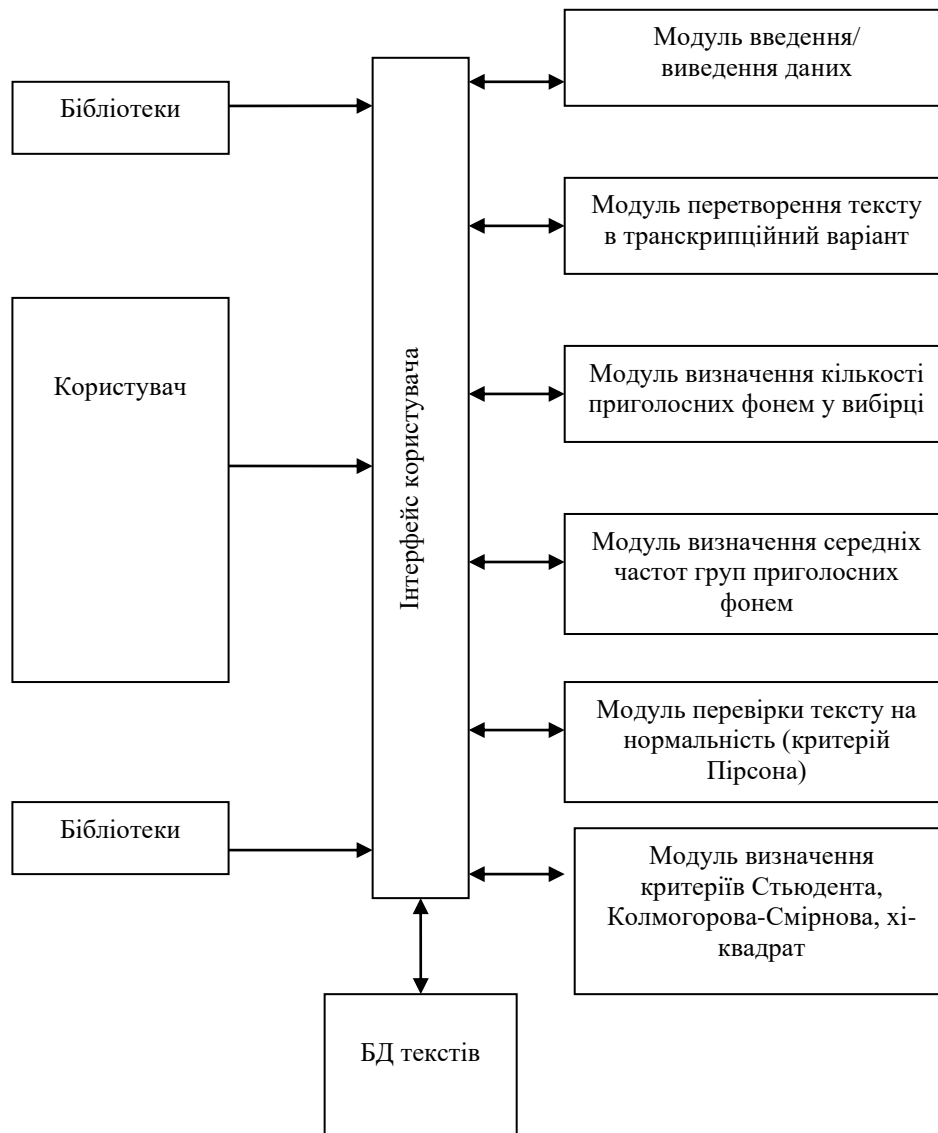


Рисунок 4.1 – Структура програмної системи диференціації фоностатистичних структур функціональних стилів англійської мови

Розроблена структура системи складається з 6 модулів.

Перший модуль структури системи призначено для введення-виведення даних вибірок із стилів, обсягом в 51 тис. фонем кожна.

Другий модуль призначено для перетворення англомовних 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 приголосних фонем закону нормального розподілу. Математичне забезпечення модуля ґрунтується на використанні критерія Пірсона.

Шостий модуль призначено для розмежування досліджуваних стилів. Математичне забезпечення модуля базується на використанні критерію Стюдента, критерію Колмогорова-Смірнова та критерію χ^2 -квадрат.

4.2 Розроблення алгоритму функціонування системи диференціації фоностатистичних структур функціональних стилів англійської мови

На основі розробленої структури системи побудовано алгоритм функціонування програми автоматизації диференціації фоностатистичних структур функціональних стилів англійської мови (рис. 4.2.).

Алгоритм функціонування системи диференціації фоностатистичних структур стилів передбачає:

Крок 1. Відбір текстів із 5 стилів і 3 підстилів та завантаження у файлах з розширенням .txt.

Крок 2. Створення транскрипційних варіантів для всіх 14 текстів стилів та підстилів.

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

Крок 4. Визначення частоти приголосних фонем у 14 вибірках.

Крок 5. Поділ вибірки кожного із 5 стилів та 3 підстилів на 51 порцію.

Крок 6. Визначення частот та середнього значення частоти вживання приголосних фонем по порціях для художнього, газетного, публіцистичного, розмовного і наукового стилів.

Крок 7. Об'єднання фонем у 8 груп за акустико-артикуляційними ознаками.



Рисунок 4.2 – Блок-схема алгоритму функціонування системи диференціації фоностатистичних структур функціональних стилів англійської мови

Крок 8. Визначення середнього значення частоти вживання групи приголосних фонем у кожному стилі та підстилі (2.1).

Крок 9. Визначення теоретичного нормального розподілу на основі емпіричного нормального розподілу (2.2, 2.3).

Крок 10. Визначення теоретичної частоти (2.5).

Крок 11. Перевірка відповідності частот вживання груп приголосних фонем закону нормального розподілу за критерієм згоди Пірсона (2.6).

Крок 12. Розмежування двох зіставлених попарно вибірок за фіксованою групою приголосних фонем за критеріями Стьюдента (2.17), Колмогорова-Смірнова (2.19) та хі-квадрат (2.20).

Розроблений алгоритм функціонування системи реалізує метод комплексного аналізу диференціації фоностатистичних структур функціональних стилів англійської мови, який дає змогу підвищити достовірність диференціації стилів.

4.3 Розроблення програмного забезпечення системи диференціації фоностатистичних структур функціональних стилів англійської мови

Програмне забезпечення диференціації текстів на фонологічному рівні, на відміну від інших рівнів мови, передбачає перетворення англійських текстів в їх транскрипційні варіанти. Програма та алгоритм диференціації текстів реалізується у таких класах (рис. 4.3):

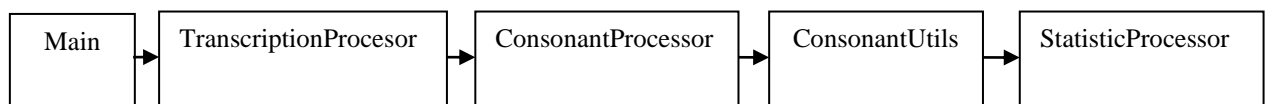


Рисунок 4.3 – Структура класів ПЗ системи диференціації фоностатистичних структур стилів

У класі Main знаходиться пакет flow, який забезпечує завантаження файлів, послідовність виконання операцій, отримання результатів обробки. Для порівняння двох текстів використано HTTP протокол, через який здійснено завантаження двох файлів і вибір критерію перевірки гіпотези на однорідність вибірок. На рис. 4.4 зображено меню вибору файлу з досліджуваним текстом та вибору критерію перевірки гіпотези на однорідність вибірок.

Key	Value	Description
<input checked="" type="checkbox"/> statistical criterion	student	
<input checked="" type="checkbox"/> file1	<input type="button" value="Вибрати файли"/> Australian Legendary Tales.txt	
<input checked="" type="checkbox"/> file2	<input type="button" value="Вибрати файли"/> Jack And The Beanstalk.txt	

Рисунок 4.4 – Меню вибору файлу з досліджуваним текстом та вибору критерію перевірки гіпотези на однорідність вибірки

Клас TranscriptionProcessor відповідає за перетворення тексту в його транскрипційний варіант. Текст розбивається на масив слів. Для кожного слова перевіряється його наявність у базі даних. У програмі розроблено POST запити двох типів: */process* і */process/transcription* [112]. Перший запит використано при відсутності транскрипційного варіанту досліджуваного тексту, другий – при його наявності. Другий запит дає змогу скоротити час роботи програми. Якщо слово відсутнє, тоді за допомогою класу від `java.net.HttpURLConnection`, що наслідується від `java.net.URLConnection`, який підтримує з'єднання по мережевому протоколу HTTP, посилається запит на сайт транскрипційного перекладу <http://upodn.com/phon.php>.

Для формування вибірки з приголосних фонем створено клас ConsonantProcessor. На рис. 4.5 зображено UML діаграму, яка описує структуру класів формування вибірки з приголосних фонем.

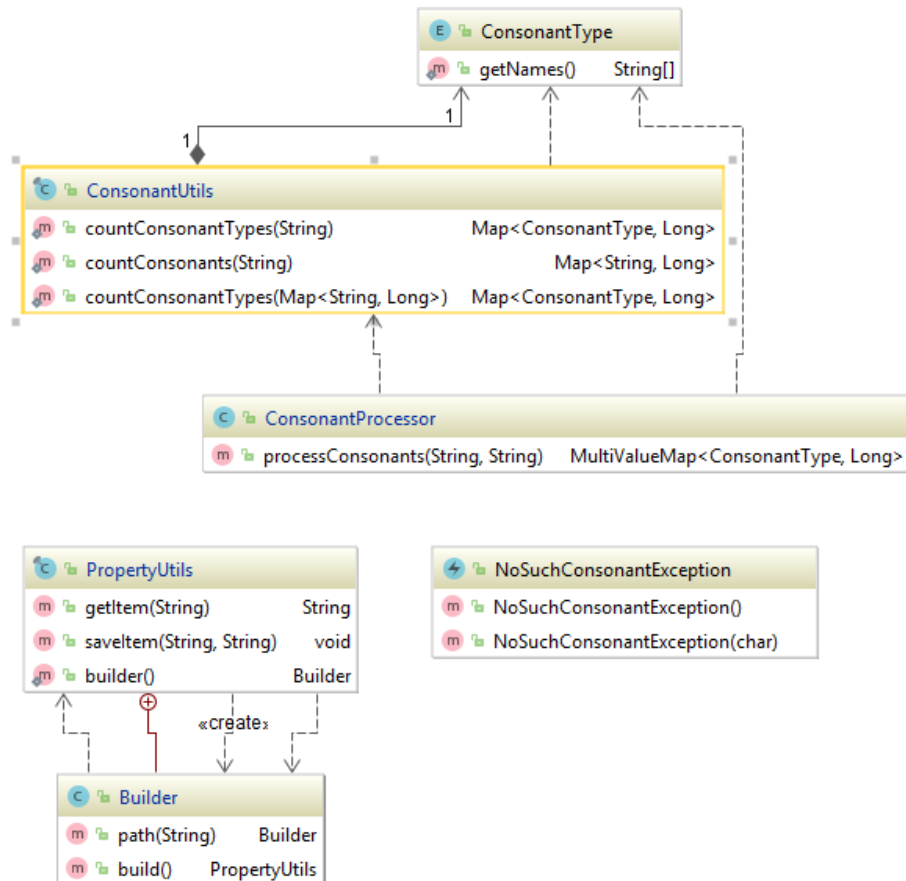


Рисунок 4.5 – Структура класів формування вибірки з приголосних фонем

Вибірка обсягом в 51000 приголосних фонем формується вилученням з транскрипційних символів всіх нефонематичних символів і голосних фонем. Якщо вибірка не має достатнього обсягу, програма повідомляє про це (Додаток Ж1). Вибірка ділиться на 51 порцію по 1000 фонем у кожній за допомогою функції *System.arraycopy()*, яка копіює створений масив символів швидше, ніж при використанні *for*-циклу. Формування порцій виконуються у циклі (Додаток Ж2). Наступним кроком є обчислення кількості приголосних фонем у кожній порції і об'єднання фонем у групи за допомогою створеного класу *ConsonantUtils* з методом *countConsonantTypes()*, який викликається для кожної групи фонем. Після проведення обчислень для кожної групи, отримано таку структуру даних: *List<Map<ConsonantType, Long>* (рис. 4.6).

```

  ▾ 0 = {HashMap@8669} size = 8
    > 0 = {HashMap$Node@8722} "LABIAL" -> "209"
    > 1 = {HashMap$Node@8723} "FRICATIVE" -> "227"
    > 2 = {HashMap$Node@8724} "DORSAL" -> "42"
    > 3 = {HashMap$Node@8725} "STOP" -> "496"
    > 4 = {HashMap$Node@8726} "VELAR" -> "59"
    > 5 = {HashMap$Node@8727} "NASAL" -> "172"
    > 6 = {HashMap$Node@8728} "SONOROUS" -> "269"
    > 7 = {HashMap$Node@8729} "CORONAL" -> "649"
  ▾ 1 = {HashMap@8670} size = 8
    > 0 = {HashMap$Node@8756} "LABIAL" -> "227"
    > 1 = {HashMap$Node@8757} "FRICATIVE" -> "256"
    > 2 = {HashMap$Node@8758} "DORSAL" -> "31"
    > 3 = {HashMap$Node@8759} "STOP" -> "481"
    > 4 = {HashMap$Node@8760} "VELAR" -> "60"
    > 5 = {HashMap$Node@8761} "NASAL" -> "158"
    > 6 = {HashMap$Node@8762} "SONOROUS" -> "258"
    > 7 = {HashMap$Node@8763} "CORONAL" -> "645"
  ▾ 2 = {HashMap@8671} size = 8
    > 0 = {HashMap$Node@8774} "LABIAL" -> "215"

```

Рисунок 4.6 – Структура даних: List<Map<ConsonantType, Long>

На основі отриманих результатів сформовано структуру даних MultiValueMap<ConsonantType, Long> (рис. 4.7). Це – структура даних типу ключ-значення, у якій ключем є тип приголосного, а значенням – вибірка.

```

"LABIAL": "209,181,215,242,228,207,195,199,202,207,222,199,206,219,227,212,215,227,215,225,248,222,206,205,1
"FRICATIVE": "212,224,217,241,244,232,235,272,253,244,234,214,240,224,226,252,266,245,222,236,221,225,261,24
"DORSAL": "66,38,50,43,24,39,36,27,29,52,38,49,26,52,43,38,39,38,50,33,36,38,55,67,29,46,35,39,45,54,47,69,5
"STOP": "493,513,495,473,441,509,490,464,488,482,499,486,483,519,490,472,474,468,489,465,502,503,443,457,476
"VELAR": "65,61,56,49,56,43,65,38,55,42,39,52,43,48,49,53,43,46,48,58,48,45,45,50,50,55,50,54,46,62,64,48,53
"NASAL": "155,177,184,180,166,192,158,162,168,164,181,172,166,182,172,162,179,153,165,147,190,164,167,158,16
"SONOROUS": "295,254,280,280,307,249,265,259,250,264,263,297,272,255,281,272,255,284,278,293,273,263,283,298
"CORONAL": "629,677,651,617,637,663,657,678,670,644,659,662,689,654,636,641,649,646,644,656,636,642,648,622,

```

Рисунок 4.7 – Структура даних MultiValueMap<ConsonantType, Long>

Процес обчислення для кожної групи фонем представлено у Додатку ЖЗ.

Наступним кроком є перевірка двох вибірок на однорідність. Якщо вибірки неоднорідні, вони відрізняються авторським стилем. Програма перевіряє вибірки на однорідність за критеріями Стюдента, χ^2 і Колмогорова-Смірнова. Для вибору критерію, необхідно у полі 'statistical criterion' вказати його назву (рис. 4.8).

	Key	Value
<input checked="" type="checkbox"/>	statistical criterion	student
<input checked="" type="checkbox"/>	file1	File ▾ <input type="button" value="Вибрати файли"/> Файл не вибрано
<input checked="" type="checkbox"/>	file2	<input type="button" value="Вибрати файли"/> Файл не вибрано
	New key	Value

Рисунок 4.8 – Меню вибору критерію перевірки гіпотези на однорідність вибірок

Клас `StatisticProcessor`, представлений на UML діаграмі, яка описує структуру класів, відповідальних за статистичне опрацювання вибірок (рис. 4.9), визначає критерій, який потрібно застосовувати. У цьому класі є метод `process`, який за параметри приймає назву критерію і всі вибірки, які попарно зіставляються.

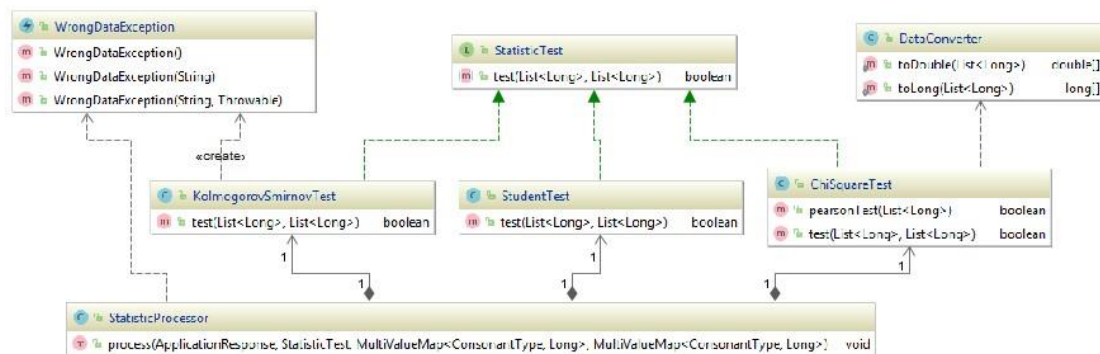


Рисунок 4.9 – Структура класів, які відповідають за статистичне опрацювання вибірки

Статистичне опрацювання вибірки здійснено за формулами наведеними у розділі 2. Процес обчислення за критерієм Пірсона (2.6) представлено у Додатку Ж4; за критерієм Стьюдента (2.17) – у Додатку Ж5; за критерієм Колмогорова-Смірнова (2.19) – у Додатку Ж6. Для обчислення критерію хі-квадрат сформовано структуру даних `List<Triple<Long, Long, Long>`. `Triple`, яка використовується для роботи з трьома величинами одночасно. У перший елемент записується значення з

першої вибірки, у другий – з другої, у третій – їхню суму. Процес обчислення за критерієм хі-квадрат (2.20) представлено у Додатку Ж7.

4.4 Розроблення інформаційного забезпечення системи диференціації фоностатистичних структур функціональних стилів англійської мови

Для забезпечення перетворення англійського тексту у транскрипційний варіант використано відкриту, кросплатформну, вбудовану базу даних H2, яка написана на мові програмування Java. База даних H2 підтримує мову SQL, має добру інтеграцію із використовуваним фреймворком Spring Boot і не потребує додаткових інсталяцій. Відповідь із сайту транскрипційного перекладу зберігається у структурі даних HashMap, яка побудована на принципі ключ-значення і дає уникнути дублікатів. Чим більше даних опрацює програма, тим менше запитів у інтернет. Із завантаженням нових текстів відбувається динамічна зміна транскрипційних знаків. У цьому разі, доцільно використовувати список `ArrayList<String>`, який є автоматичним розширювальним списком, і за допомогою якого зручно виконувати операції з динамічними даними. Приклад використаної спискової структури даних наведено на рис. 4.10.

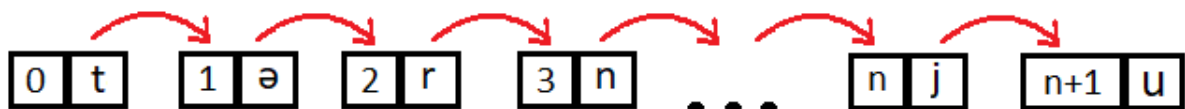


Рисунок 4.10 – Приклад спискової структури даних

Завдяки спискам кожне слово нового тексту зіставлено із уже існуючими словами у програмі. Якщо є збіг слів, то в програмі передається команда, що слово з індексом «х» потребує транскрипційного варіанта. Тоді список дає стрічку з перетворенням у транскрипційні знаки з відповідним індексом. Якщо збігу немає, тоді до списку записується символ з тим самим індексом, що і його перетворення у списку.

4.5 Результати дослідження фоностатистичної структури розмовного стилю

У дослідженні розглядається той різновид розмовного стилю, який є характерним для повсякденного спілкування. Цей різновид функціонує у сфері побуту, поза виробничою та суспільно-політичною діяльністю. Вибірку представляють тексти із розмовника, в яких домінує сучасне діалогічне мовлення [113].

Вибірка є, у значній мірі, позбавлена елементів експресивності і емоційності, які є характерними для фамільярно-розмовного різновиду розмовного стилю і просторіччя. Тому досліджуваний літературно-розмовний різновид розмовного стилю є відносно немаркованим спеціальними термінами, що вживаються у певних сферах людської діяльності, а також експресивними та емоційними елементами. Слід заважити, що досліджуваний різновид формує серцевину основного словникового фонду мови, тобто представляє відомі носіям слова, які вживаються у всіх стилевих різновидах.

Перш ніж представити результати попарного порівняння розмовного, художнього, газетного і наукового стилів, розглянемо специфіку діалогічного мовлення досліджуваного різновиду розмовного стилю. Характеризуючи англійське діалогічне мовлення, слід зазначити перевагу повторів. На відміну від українського мовлення, в англійському діалогічному мовленні набагато рідше вживаються репліки з повтором, ускладненим додатковими лексико-граматичними елементами.

Розглядаючи характерні особливості драми Б. Шоу, неважко помітити, що, незважаючи на літературне опрацювання усного мовлення, діалог відображає основні закономірності живого розмовного мовлення. Тому невипадково у сучасному мовознавстві підкреслюється зближення літературної мови і розмовної мови у межах художнього твору. Зокрема, Й. Бароуз відзначає, що немала кількість творів написана, у значній мірі, розмовним стилем [114].

Слід зазначити, що елементи розмовного стилю потрапляють у художній твір не тільки через свою природну підвищену здатність до проникнення, але і як засоби типізації персонажів, які автор вживає з метою відтворення реального спілкування.

Взаємодія статистичних структур розмовного стилю з художнім, газетним і науковим стилями, одержана за методом гіпотез, є представлена у таблицях А.1 – А.18 Додатку А; за методом ранжування у таблицях Е.1, Е.2, Е.3 Додатку Е [115]; за методом визначення відстаней між стилями у таблицях Е.4, Е.5, Е.6 Додатку Е.

Отримані методом гіпотез дані вказують на істотні відмінності між розмовним стилем та наступними функціональними стилями: художнім (поезія, художня проза, драма), науковим і газетним. Аналіз отриманих даних дає змогу зробити наступні припущення і висновки. Тотожні результати отримано у порівнянні: розмовний стиль – поезія Дж. Г. Байрона; розмовний стиль – поезія Т. Мура. В обох випадках розмовний стиль істотно відрізняється від поезії за групами передньоязикових, середньоязикових, задньоязикових, носових, щілинних та зімкнених приголосних при не врахуванні позиції фонемі в слові. Ці результати підтверджують близькість текстів поезії, як належних до одного стилю, з одного боку, а також наявність спільних рис у творчості Дж. Г. Байрона і Т. Мура. Дж. Г. Байрон і Т. Мур є представниками романтизму. Добре відомі їх творчі стосунки та вплив творчості Дж. Г. Байрона на творчість Т. Мура [116].

Розглядаючи відмінність поезії Байрона від розмовного стилю, слід зауважити, що у досліджених текстах простежується наближення до усного мовлення, яке дає змогу простіше і легше висловлювати ліричні емоції. Варто зазначити, що Байрон як прихильник поезії серйозного та піднесеного стилю, черпав стилістичні засоби з джерел античної міфології. Тому в його поезії, поряд із простими повсякденними словами, зустрічаються і архаїзми, і діалектизми. Саме наявність цих специфічних для його поезії елементів, як і особливостей, властивих поезії взагалі, відображено в істотній різниці зіставлених числових показників поезії Байрона з показниками розмовного стилю [117].

Враховуючи позицію фонемі в слові, отримано майже тотожні розходження між розмовним стилем і поезією Байрона та між розмовним стилем і поезією Мура.

Числові дані представлено у таблицях А.1 – А.6 Додатку А для всіх трьох випадків позиції фонем в слові.

Художня проза Дж. Г. Байрона істотно відрізняється від розмовного стилю за наступними групами: губних, передньоязикових, середньоязикових, задньоязикових, сонорних, щілинних і зімкнених для випадку не врахування позиції фонем в слові. Ці розходження, свідчать про використання Байроном різноманітних стилістичних експресивних засобів, які властиві художній прозі [118].

Враховуючи позицію фонем в слові, отримуємо числове підтвердження відмінності між текстом, в якому домінують емоційні та експресивні елементи, і текстом, в якому переважає побутово-розмовна лексика. Так, для випадку позиції фонем в кінці слова, стилі істотно відрізняються за сімома групами фонем, а для випадку позиції фонем на початку слова – за п'ятьма.

Дані для всіх трьох випадків представлено в таблицях А.7 – А.9 Додатку А.

Важливими результатами дослідження є дані, отримані при порівнянні розмовного стилю з драмою. На відміну від попередніх розглянутих стилів, драма у найменшій мірі, відрізняється від розмовного стилю. Так, істотні розходження встановлено лише за чотирма групами фонем для всіх трьох випадків. Аналізуючи кожен випадок зокрема, слід зазначити наступні розходження: не враховуючи позицію фонем в слові, стилі відрізняються за групами губних, передньоязикових, носових і зімкнених; для випадку позиції фонем на початку слова – за групами губних, середньоязикових, носових і зімкнених; для випадку позиції фонем в кінці слова – за групами передньоязикових, задньоязикових, сонорних і зімкнених [119].

Таким чином, група зімкнених фонем відзначається найбільшою стилерозрізняльною здатністю, бо за нею виявлено істотні розмежування порівнюваних стилів для всіх трьох випадків [120, 121].

Зазначена близькість пояснюється переважанням діалогічної форми мовлення в обох стилях та особливостями творчої манери викладу Б. Шоу. Правдиве відтворення дійсності відображено у щоденно-побутовому змісті реплік персонажів драми. Елементи ірландського просторіччя та художньо-експресивні засоби спричинюють істотні розходження за вищеназваними групами фонем.

У таблицях А.10 – А.12 Додатку А представлено дані для досліджених нами випадків. За цими даними побудовано модель (рис. 4.11).

За методом ранжування визначено близькість (різниця рангових показників – 1) за групами передньоязикових, середньоязикових, сонорних для випадку позиції фонему на початку слова. Результати одержані за методом визначення відстаней між стилями наведено у таблицях Е.4 – Е.6 Додатку Е. Порівняння газетного стилю з розмовним показує істотні розходження за шістьма групами фонем для випадків не врахування позиції фонему в слові та позиції фонему на початку слова. Це відповідно групи передньоязикових, середньоязикових, задньоязикових, носових, щілинних і зімкнених; губних, передньоязикових, середньоязикових, задньоязикових, сонорних і зімкнених. Ці розходження можна пояснити значною кількістю слів, які вживаються у різних сферах людської діяльності – політичної, економічної, культурної та спортивної. Для меншої кількості груп приголосних фонем істотні розходження встановлено у випадку позиції фонему в кінці слова. Стилі істотно відрізняються лише за групами задньоязикових, щілинних і зімкнених [122, 123]. Зазначені розходження зумовлені характерними для газетного стилю лексичними одиницями, фонемний склад яких у даній позиції є, в більшій мірі, ніж в інших позиціях, тотожний фонемному складу лексичних одиниць розмовного стилю. Дані представлено в таблицях А.13 – А.15 Додатку А. Науковий стиль істотно відрізняється від розмовного стилю за групами губних, передньоязикових, середньоязикових, задньоязикових, носових, щілинних і зімкнених для випадку не врахування позиції фонему в слові. Для випадку позиції фонему на початку слова істотні розходження виявлено за групами губних, передньоязикових, середньоязикових, задньоязикових, носових, сонорних і зімкнених. Отже, істотні розходження встановлено за більшістю – сімома групами фонем для зазначених двох випадків. Таке чітко виражене розходження порівнюваних стилів пояснюється тим, що насиченість досліджуваних технічних текстів спеціальними термінами, вживаними у галузі технічної фізики, природно, зумовлює відмінність від розмовного стилю, яка є підтверджена статистичними даними [124].

Менш чітке розмежування стилів отримано для випадку позиції фонемі в кінці слова. Стили відрізняються за чотирма групами фонем. Це групи губних, задньоязикових, щілинних і зімкнених. Очевидно, що звукова сторона лексичних одиниць порівнюваних стилів відрізняється, у меншій мірі, для даного випадку, ніж для двох попередніх. У таблицях А.16–А.18 Додатку А представлено дані для всіх трьох випадків. Результати диференціації художнього (поезія, художня проза, драматургія), розмовного, газетного та наукового стилів наведено у таблицях 4.1, 4.2, 4.3.

Удосконалено статистичну модель стилеві диференціації текстів, яка репрезентує встановлені істотні відмінності функціональних стилів англійської мови на фонологічному рівні з довірчою ймовірністю 0,95 (рис. 4.11).

Таблиця 4.1

Позиція фонемі в слові неврахована

Стиль	ПзБ	ПзМ	Пр	Др	ГС	НС	РС
	І	І	І	І	І	І	І
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	-

Статистична модель дає змогу підвищити достовірність отриманих результатів за кількістю груп приголосних фонем, за якими вибірки відрізняються істотно. Модель визначає приналежність досліджуваного тексту до розмовного стилю [125 – 127].

Таблиця 4.2

Фонема на початку слова

Стиль	ПзБ	ПзМ	Пр	Др	ГС	НС	РС
	П	П	П	П	П	П	П
ПзБ	-	2	8	6	7	8	6
ПзМ	2	-	5	7	7	6	5
Пр	8	5	-	6	3	6	5
Др	6	7	6	-	4	7	4
ГС	7	7	3	4	-	6	6
НС	8	6	6	7	6	-	7
РС	6	5	5	4	6	7	-

Таблиця 4.3

Фонема в кінці слова

Стиль	ПзБ	ПзМ	Пр	Др	ГС	НС	РС
	Ш	Ш	Ш	Ш	Ш	Ш	Ш
ПзБ	-	2	5	5	4	6	4
ПзМ	2	-	5	2	4	5	4
Пр	5	5	-	6	4	6	7
Др	5	2	6	-	5	6	4
ГС	4	4	4	5	-	4	3
НС	6	5	6	6	4	-	4
РС	4	4	7	4	3	4	-

У таблицях 4.1, 4.2, 4.3 використано такі позначення: ПзБ – поезія Дж.Г. Байрона, ПзМ – поезія Т. Мура, Пр – художня проза Дж. Г. Байрона, Др – драматургія Б. Шоу, РС – розмовний стиль, ГС – газетний стиль, НС – науковий стиль.

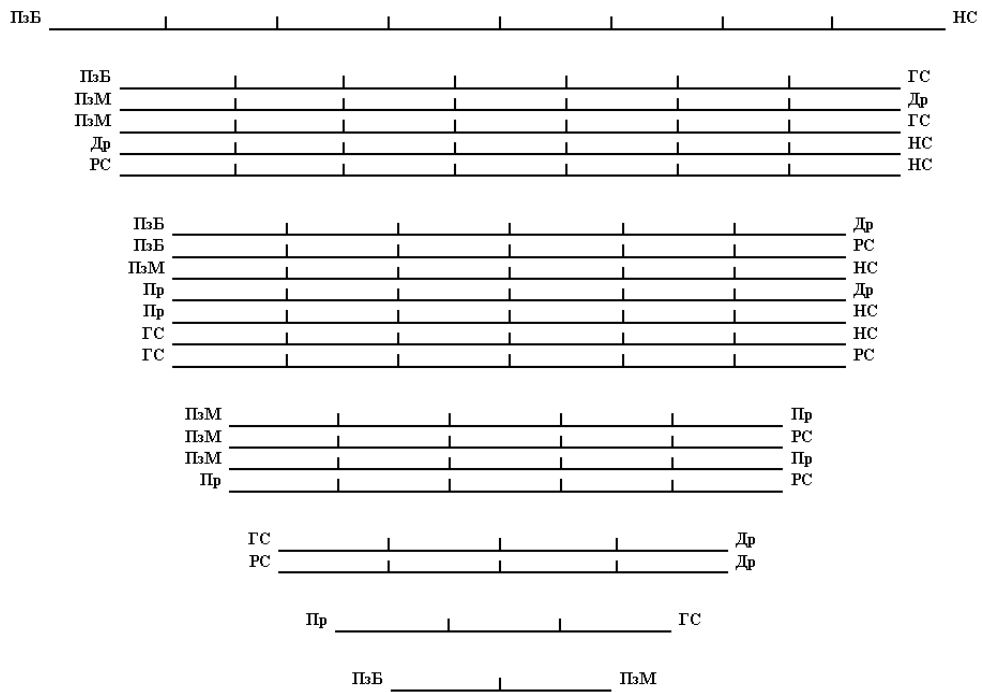


Рисунок 4.11 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі диференціації художнього (поезія, художня проза, драматургія), розмовного, газетного та наукового стилів на фонологічному рівні

4.6 Результати дослідження фоностатистичної структури художнього стилю

4.6.1 Результати дослідження фоностатистичної структури поезії

Вірш став віршем у результаті дії і організації певних аспектів звукової системи. Їхня організація зумовлює сильніший зв'язок слів в поезії, ніж у прозі. Вона визначається високою гармонійністю і злагодженістю. Очевидним є те, що строга упорядкованість поезії зумовлює вживання лексики, яка створює мелодійний естетичний ефект. Чи потрібна у даному випадку якась особлива лексика, вживана лише в поезії, чи поет використовує загальноживану лексику даної мовної системи? Це питання було особливо актуальним в епохи класицизму та романтизму. З метою досягнення особливої вишуканості поетичні засоби вживались відповідно

до строгих канонів. Ефект поетичної вишуканості створювався штучно, завдяки добору слів, які через особливе звучання, створювали поетичний ефект, який відповідав авторському задуму. Зазвичай, такі новостворені автором лексеми не відзначалися високою частотою вживання.

Проте поети-романтики не дотримувались цього напряму і намагались збагатити мову нації лексикою з різних джерел. Творам представника романтизму Дж. Г. Байрона властиві і архаїчні, і класичні, і фольклорні елементи.

Таким чином, досліджувані нами тексти поезії не відзначаються високою частотою вживання поетизмів, тобто слів, які несуть поетичну конотацію лише у поетичних контекстах. Окремі поетизми є представленими історизмами, архаїзмами, діалектизмами, які з'явилися в епоху класицизму.

Які ж конкретні мовні засоби використовує Байрон в досліджених творах? Розглянемо докладніше ці твори.

Ліро-епічна поема "The Corsair" (1814) присвячена Томасу Муру [128]. Високу частоту вживання мають лексичні одиниці, які є характерними для поезії того часу взагалі. Це вживання 'tis, 'twas, thou, thee, thy, thine, er, whate'er, hath, methinks' (methought). Проте перелічені слова не є чимось особливим, притаманним лише поезії. Вони можуть вживатися і в усному мовленні.

Поема "The Prisoner of Chillon" (1816) написана Байроном у швейцарський період його творчості [129]. Слід наголосити на високій частоті вживання архаїзмів у поезії того часу. Серед них слова, які поступово виходять із загального вжитку – thou, thee, thy, thine і hath, а також слова, які вже не вживаються, але все ще розпізнаються носіями мови - methinks (methought) [129]. Вживання архаїзмів в поезії вважається основною і, чи не єдиною, розрізняльною ознакою романтичного стилю і пояснюється прагненням романтиків відродити стародавню літературу.

Байрон використовує лексику розмовного стилю, діалектизми та архаїзми. Взаємодія фоностатистичних структур поезії Байрона з іншими досліджуваними стилями, одержана за методом гіпотез, є представленою у таблицях А.19 – А.30 Додатку А; за методом ранжування – у таблицях Е.1 – Е.3 Додатку Е; за методом визначення відстаней між стилями – у таблицях Е.4 – Е.6 Додатку Е.

Про специфіку поезії Байрона на фонологічному рівні можна зробити висновки за отриманим у дослідженнях розподілом частот приголосних фонем. При порівнянні поезії Байрона з його художньою прозою виявлено істотні розходження за всіма розглянутими групами фонем для випадку позиції фонем на початку слова. Це важливі результати, бо вони є надійним доказом того, що поезія Байрона залишається поезією у найвищому розумінні цього слова, незважаючи на вплив елементів розмовного стилю. Така чітка відмінність поезії і художньої прози Байрона свідчить про більш-менш чітку окресленість даних підстилів художнього стилю Байрона для даного випадку позиції фонем в слові.

Для випадку фонем у кінці слова відмінність менш різка. Істотні відмінності виявлено для п'яти із семи груп фонем. Це групи: губних, передньоязикових, задньоязикових, сонорних і зімкнених. У даній позиції спостережено взаємопроникнення елементів порівнюваних підстилів, яке пояснюється спільними рисами поезії та художньої прози – тобто наявністю елементів експресивності та емоційності.

Про ще вищий ступінь взаємопроникнення можна говорити, не враховуючи позицію фонем в слові, коли отримуємо істотні відмінності ще для меншої кількості груп – чотирьох із восьми: губних, передньоязикових, сонорних, щілинних. Таким чином, специфіка фонемного складу лексичних одиниць, представлена у значеннях середніх частот груп приголосних фонем для кожного із досліджуваних випадків, визначає істотні розходження між порівнюваними вибірками та дає інформацію про ступінь віддаленості чи близькості між ними для кожного конкретного випадку. Вищезгадані розходження наведено у таблицях А.19 – А.21 Додатку А.

Приблизно однакові результати отримано для всіх трьох досліджуваних випадків при порівнянні поезії Дж. Г. Байрона з драмою Б. Шоу. У двох випадках порівнювані підстилі істотно відрізняються за шістьма групами фонем, а в одному – за п'ятьма. Проте для кожного окремого випадку, підстилі відрізняються за іншими групами фонем. Так, не враховуючи позицію фонем в слові, істотні відмінності встановлено для груп губних, передньоязикових, середньоязикових, носових,

щілинних, сонорних; для випадку позиції фонему на початку слова – це групи губних, передньоязикових, середньоязикових, щілинних, зімкнених, сонорних і для випадку позиції фонему в кінці слова – це групи губних, передньоязикових, задньоязикових, сонорних, щілинних. Розходження за більшістю груп фонем свідчать про природну відмінність поетичного тексту Байрона від прозового тексту драми Шоу. Дані таблиць А.22 – А.24 Додатку А підтверджують вищесказане.

У порівнянні поезії Байрона з газетним стилем одержано підтвердження відмінності. Позиція фонему на початку слова виявилась найбільш вдалою для встановлення істотних розходжень. У цій позиції вибірки істотно відрізняються за сімома із восьми групами фонем. Для меншої кількості груп приголосних фонем отримано істотні розходження у позиції фонему в кінці слова. Це групи передньоязикових, задньоязикових, щілинних і зімкнених. Проте найвищий ступінь схожості виявлено у випадку не врахування позиції фонему у слові для однієї групи фонем – передньоязикових.

Розходження між даними стилями представлено у таблицях А.25 – А.27 Додатку А. Відмінність поезії Байрона від наукового стилю підтверджена у числовому вираженні для позиції фонему на початку слова, у якій істотні розходження встановлено для всіх восьми груп приголосних фонем. Наявність значної кількості специфічних термінів, які обслуговують певну наукову сферу, є тим фактором, який відмежовує науковий стиль від поезії на лексичному рівні. Майже ідентична картина спостерігається для позиції фонему у кінці слова. Лише для групи сонорних розходження є випадковими. Не враховуючи позицію фонему в слові, отримано протилежні результати. Істотні розходження отримано лише для двох груп приголосних фонем: губних та зімкнених. У даному випадку спостережено схожість фонематичного наповнення. Порівняння поезії Байрона з науковим стилем представлено у таблицях А.28 – А.30 Додатку А.

При порівнянні поезії Байрона з розмовним стилем істотні розходження виявлено для більшості груп приголосних фонем в випадках не врахування позиції фонему в слові та позиції фонему на початку слова. Для випадку позиції фонему в

кінці слова вибірки відрізняються за чотирма групами фонем. Результати досліджень наведено у в таблицях А.1 – А.3 Додатку А.

Другою вибіркою з підстилю поезії є вибірка з творчості Томаса Мура. Томас Мур справедливо вважається першим поетом Ірландії. В основу його творчості лягли естетичні ідеї романтизму, проголошені У. Вордсвортом. Саме У. Вордсворт та С. Т Кольридж сформулювали один з основних принципів нової поезії: "Поезія для всіх і мова доступна кожному" [130]. Мур є поетом-романтиком, джерелом творчості якого є народна творчість, народна пісня.

У поезії Т. Мура спостережено перехід елементів розмовного стилю в поезію. У цьому зв'язку слушно зауважити, що надання поезії національного колориту було важливою рисою творчості поетів-романтиків. Лірична та музикальна поезія Мура тісно пов'язана з народною творчістю, і тому Байрон називає Мура 'Бернсом Ірландії' [131].

У дослідженні об'єктом статистичного обстеження є друга і четверта поеми поетичного твору "Lalla Rookh" (Лалла Рук) [132]. За жанром це роман у віршах. Мур працював над цим твором з 1812 по 1817 рік. Аналізуючи поему в лексичному плані, слід зазначити вживання архаїзмів *thou, thee, thy і thine*; скорочених форм *'t is, 't was, e'er, o'er, where'er, ne'er, thro', tho'*; поетичних форм *morn, ye* та ін.

"The Light of the Haram" (Світло гарему) – четверта поема твору "Лалла Рук" є своєрідним апофеозом краси. У цій поемі частіше, ніж у попередній, вживаються власні імена та назви місцевостей, взяті з Біблії, міфології. Аналізуючи склад лексичного рівня, слід додати, що крім слів та скорочених форм, вжитих у поемах "Пері і Ангел" та "Світло гарему", вживаються такі дієслівні форми: *hath, hast, dost, banisht, vanisht, husht, pluckt, maskt, fixt* та ін [133].

Взаємодія фоностатистичних структур поезії Мура з іншими досліджуваними стилями, одержана за методом гіпотез, є представленою у таблицях А.31 – А.42 Додатку А; за методом ранжування – у таблицях Е.1 – Е.3 Додатку Е; за методом визначення відстаней між стилями – у таблицях Е.4 – Е.6 Додатку Е. Специфіка фонологічного рівня поезії Т. Мура у дослідженні визначається у порівнянні числових даних частоти вживання фонем підстилю поезії з числовими даними

інших досліджуваних стилів. Так, при порівнянні поезії Т. Мура з художньою прозою Дж. Г. Байрона істотні розходження виявлено для п'яти груп фонем, враховуючи позицію фонем в слові. Для позиції фонем на початку слова – це групи губних, передньоязикових, середньоязикових, сонорних і щілинних. Для позиції фонем в кінці слова – це групи передньоязикових, задньоязикових, носових, сонорних і зімкнених. Таким чином, виявлена деяка відмінність і деяка подібність [134]. Відмінність, очевидно, пов'язана з високою мелодійністю та ліризмом поезії Мура, що, зрозуміло, не може характеризувати прозового тексту завважень Байрона. Подібність впливає з наявності в обох підстилях загальноновживаних слів та схожості фонематичного наповнення. Крім того, подібність є більшою при не врахуванні позиції фонем в слові. Істотні розходження визначено за трьома групами приголосних фонем: передньоязиковими, сонорними і зімкненими. Числовий матеріал для всіх трьох позицій представлено у таблицях А.31 – А.33 Додатку А.

Деяко суперечливі результати отримано при порівнянні поезії Т. Мура з драмою Б. Шоу. Істотні розходження встановлено для груп губних, передньоязикових, середньоязикових, задньоязикових, носових, щілинних і зімкнених, не враховуючи позицію фонем в слові, та для груп губних, передньоязикових, середньоязикових, носових, сонорних, щілинних і зімкнених для позиції фонем на початку слова. Зовсім іншого характеру отримано результати для позиції фонем у кінці слова. У цьому випадку підстилі відрізняються лише за групами передньоязикових і щілинних [135]. Якщо порівняти розходження між поезією Байрона і поезією Мура, з одного боку, і поезією Мура з драмою Шоу, з другого боку, то, очевидно, що у першому випадку творчість двох поетів повинна мати більше спільних елементів, ніж творчість поета і драматурга. Підтвердження цього припущення отримано за результатами статистичного обстеження для двох перших вищезгаданих випадків. Випадок позиції фонем в кінці слова виявився найменш вдалим для встановлення істотних розходжень між двома підстилями художнього стилю, оскільки у цій позиції, групи фонем мають досить малу

стилерозрізняльну здатність, яка є зумовлена схожістю фонематичного наповнення. Усі досліджувані випадки представлено у таблицях А.34 – А.36 Додатку А.

Істотні розходження, які виявлено при порівнянні поезії Мура з газетним стилем, підтверджують наявність об'єктивних відмінностей між ними. Більшість груп фонем отримує показники, за якими поезія і газетний стиль істотно відрізняються без врахування позиції фонемати в слові та в позиції фонемати на початку слова [136]. У першому випадку – це групи губних, передньоязикових, задньоязикових, носових, щілинних і зімкнених. У другому випадку – це групи губних, передньоязикових, середньоязикових, задньоязикових, носових, сонорних і щілинних. Не так чітко виражена відмінність між поезією і газетним стилем для позиції фонемати в кінці слова. У цьому випадку вибірки істотно відрізняються лише за чотирма групами фонем. Це групи губних, передньоязикових, щілинних і зімкнених. Отже, у цьому випадку спостережено схожість фонематичного наповнення. Таблиці А.37 – А.39 Додатку А наводять дані розходжень між поезією і газетним стилем для розглянутих випадків.

Приблизно однакова картина отримана при порівнянні поезії Т.Мура з науковим стилем для всіх випадків. Не враховуючи позицію фонемати в слові, істотні розходження встановлено для груп губних, середньоязикових, носових, сонорних, щілинних і зімкнених. Для позиції фонемати на початку слова – це групи губних, передньоязикових, задньоязикових, носових, сонорних і щілинних. Для позиції фонемати в кінці слова стилерозрізняльну роль відіграють групи губних, передньоязикових, задньоязикових, щілинних і зімкнених [137, 138]. Отже, отримано важливі результати, які підтверджують об'єктивну відмінність технічного тексту з значною кількістю специфічних для даної галузі термінів від мелодійної поезії Т. Мура. У даному порівнянні стилів спостережено схожість фонематичного наповнення, бо не за всіма групами фонем встановлено істотні відмінності. Таблиці А.40–А.42 Додатку А представляють дані відмінностей між стилями для трьох розглянутих випадків.

Аналіз результатів порівняння поезії Т. Мура з розмовним стилем представлено у розділі 4.2. Варто нагадати, що при порівнянні даних стилів, істотні

розходження встановлено для більшості груп фонем у випадках не врахування позиції фонем в слові та позиції фонем на початку слова. За чотирма групами фонем стилі істотно відрізняються для випадку позиції фонем в кінці слова. (Таблиці А.4 – А.6 Додатку А).

Показові результати отримано при зіставленні поезії Дж. Г. Байрона з поезією Т. Мура. Істотні розходження встановлено лише за двома групами фонем, враховуючи позицію фонем в слові. Так, для позиції фонем на початку слова – це групи задньоязикових і зімкнених, а для позиції фонем в кінці слова – групи губних і щілинних. Для випадку не врахування позиції фонем в слові встановлено істотні відмінності за групами сонорних, щілинних і зімкнених. Отже, у даному випадку можна говорити про близькість поезії двох різних авторів на фонологічному рівні. Слід зазначити, що встановлена близькість не випадкова, і підтверджує чимало фактів, які свідчать про творчу близькість поетів. Це, перш за все, приналежність поетів до одного напрямку – романтизму, звертання до міфології, біблійної тематики, що на лексичному рівні зумовило вживання архаїзмів: *thou, thee, thy, thine* та *in* [139 – 142]. Обидва автори, відкидаючи традиційні рамки класицизму, писали поезію живою розмовною мовою, вживаючи діалектизми і просторіччя. Відома творча співпраця поетів. Беручи до уваги належність Байрона і Мура до одного історичного періоду, одного творчого напрямку та їх творчу співпрацю, можна вважати, що це ті фактори, які зумовили вищезгадані спільні особливості поезії. Числові дані представлено у таблицях А.43 – А.45 Додатку А.

Художня проза Дж. Г. Байрона істотно відрізняється від наукового стилю за шістьма групами фонем для всіх трьох випадків. Для випадку не врахування позиції фонем в слові – це групи губних, передньоязикових, носових, сонорних, щілинних і зімкнених. Для випадку позиції фонем на початку слова – це групи губних, передньоязикових, задньоязикових, носових, сонорних і щілинних. Для випадку позиції фонем в кінці слова – це групи губних, передньоязикових, носових, сонорних, щілинних і зімкнених. Значна кількість спеціальних термінів, які вживаються у галузі технічної фізики, зумовила істотні відмінності за більшістю

груп приголосних фонем наукового стилю від художньої прози Дж. Г. Байрона. Таблиці А.52 – А.54 Додатку А представляють дані для всіх трьох випадків.

Дані наведених таблиць та подана нижче вдосконалена статистична модель авторської диференціації текстів (рис. 4.12) підтверджують, встановлену за різницею середніх частот, відносну близькість поезії Дж. Г. Байрона і Т. Мура (істотні відмінності визначено з довірчою ймовірністю 0,95 лише за трьома групами фонем при не врахуванні позиції фонем в слові і за двома групами фонем для позицій фонем на початку і в кінці слова), яка є зумовлена приналежністю поетів до одного історичного періоду і літературного напрямку, а також їх творчою співпрацею, впливом творчих засад Дж. Г. Байрона на творчість Т. Мура.

Статистична модель з більшою достовірністю уможливило встановлення авторства досліджуваного тексту за кількістю груп приголосних фонем, за якими вибірки відрізняються істотно. Зіставлені вибірки належать до одного стилю, підстилю, одного історичного періоду та літературного напрямку [143].

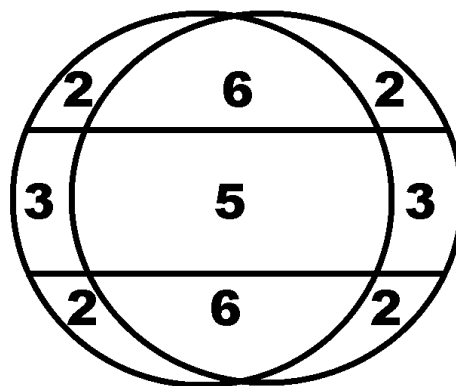


Рисунок 4.12 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі авторської диференціації текстів поезії Дж. Г. Байрона і Т. Мура

На рис. 4.12 використано такі позначення: 2, 3 – невелика кількість груп фонем, за якими встановлено істотні відмінності (тексти схожі); 5, 6 – велика кількість груп фонем, за якими встановлено істотні відмінності (тексти відмінні).

4.6.2 Результати дослідження фоностатистичної структури художньої прози

Художня проза у дослідженні розглядається на матеріалі творчості Дж. Г. Байрона. Об'єктом дослідження є прозові зауваги до пісень поеми "Childe Harold's Pilgrimage [144].

Відмінність художньої прози як підстилю художнього стилю (стилю художньої літератури) від інших функціональних стилів широко дискутується на сучасному етапі розвитку стилістики. Складність цього завдання пояснюється специфікою художньої прози. Письменник, реалістично зображаючи персонажа, передає всі ті деталі, які можуть його охарактеризувати. Наприклад, лексика фамільярно-розмовного стилю вживається для характеристики малоосвідчених людей, спеціальна лексика вживається, якщо персонажами є люди певних професій, поезія стає невід'ємною частиною у творі, присвяченому поетові. Таким чином, художня проза не є чимось однорідним, і чітко відмежованим від інших функціональних стилів, а є своєрідним поєднанням різних стилів. Слід зазначити, що художня проза і драма, у більшій мірі, зазнають проникнення елементів з інших функціональних стилів, ніж поезія, хоча всі три є підстилями художнього стилю.

Розглядаючи проблему виокремлення стилю художньої літератури в окремий стиль, М. Н. Кожина вважає, що елементи, які проникають у стиль художньої літератури з інших стилів, не відтворюють особливостей цих стилів в буквальному розумінні, а піддаються художньому "перетворенню" [64]. Диференційною ознакою стилю художньої літератури, а отже, і підстилю художньої прози як його складової частини, є підвищена експресія і емоційність.

Розглядаючи підстиль стилю художньої літератури – художню прозу на вищезазначеному матеріалі, слід зазначити, що у віршах представлені сентиментальні розважання, у прозі – виклад конкретних подій.

Прозові зауваги до поеми Байрона "Паломництво Чайльд Гарольда" представлені коментуванням тієї чи іншої історичної події, роздумами автора над політичним станом країни, описом тієї чи іншої місцевості, викладенням історії

певної пам'ятки старовини, окремими епізодами з міфології. На лексичному рівні зазначений зміст завваг передається лексичними одиницями, які вживаються у суспільно-політичній сфері, сфері історії та географії. Проте проникнення елементів характерних для інших функціональних стилів не позбавляє тексту характерних особливостей художньої прози. Вживаючись у даному контексті, 'чужі' елементи, піддаються художньому опрацюванню, набуваючи експресивного забарвлення [145].

Взаємодія фоностатистичних структур художньої прози Байрона з підстилями поезії і драми та розмовним, газетним і науковим стилями одержана за методом гіпотез і представлена у наступних таблицях: А.7 – А.9 Додатку А; А.19 – А.21; А.31 – А.33 Додатку А; А.46 – А.54 Додатку А; за методом ранжування – у таблицях Е.1 – Е.3 Додатку Е; за методом визначення відстаней між стилями – у таблицях Е.4 – Е.6 Додатку Е. У попередніх підрозділах (4.5; 4.6.1) всебічно висвітлено відмінність художньої прози від поезії Байрона, Мура та від розмовного стилю. У даному підрозділі наведено короткий виклад результатів зазначених порівнянь.

Порівнюючи художню прозу і поезію Дж. Г. Байрона, отримано неоднакові результати для всіх трьох випадків позиції фонем в слові. Відрізняються дані для випадків не врахування позиції фонем в слові та позиції фонем на початку слова. Істотні розходження встановлено у першому випадку за чотирма, а удругому за восьмома групами фонем. Для випадку позиції фонем в кінці слова підстилі відрізняються за п'ятьма групами фонем. Зазначений різнобій отриманих даних зумовлений об'єктивною відмінністю художньої прози і поезії. Дія вищезгаданих факторів підтверджується виявленням різної кількості груп приголосних фонем, за якими встановлено істотні розходження для трьох досліджуваних випадків (Таблиці А.19 – А.21 Додатку А).

Поезія Мура із властивою їй мелодійністю і співучістю та тексти прозових завваг Дж. Г. Байрона, за даними числових показників середньої частоти вживання груп фонем істотно відрізняються за п'ятьма групами фонем для випадків позиції фонем на початку слова та позиції фонем в кінці слова. Не враховуючи позицію

фонемі в слові, отримано істотні розходження за трьома групами фонем (Таблиці А.31 – А.33 Додатку А).

Художня проза Байрона істотно відрізняється від розмовного стилю за сімома групами фонем для випадків не врахування позиції фонемі в слові та позиції фонемі в кінці слова. Для випадку позиції фонемі на початку слова вибірки істотно різняться за п'ятьма групами фонем. Встановлення істотних розходжень за більшістю груп фонем підтверджує той факт, що, не зважаючи на проникнення елементів стилю розмовної мови у художню прозу, художня проза, з притаманною їй експресивною забарвленістю та здатністю поглинати елементи інших стилів, займає своє окреме місце у системі функціональних стилів (Таблиці А.7 – А.9 Додатку А).

У даному підрозділі всебічно розглядаються порівняння художньої прози Байрона з драмою Шоу, з газетним та науковим стилями.

Проза Байрона істотно відрізняється від драми Шоу за сімома групами приголосних фонем для випадку не врахування позиції фонемі в слові. Це групи губних, передньоязикових, середньоязикових, задньоязикових, носових, щілинних і сонорних. Для випадків позиції фонемі на початку та в кінці слова істотні відмінності встановлено за шістьма групами приголосних фонем. Це відповідно групи: губних, середньоязикових, носових, сонорних, щілинних і зімкнених; передньоязикових, задньоязикових, носових, сонорних, щілинних і зімкнених [146].

Основним фактором, який зумовив істотні розходження за більшістю груп фонем, є значна кількість реплік ірландською говіркою у досліджуваній драмі. У підрозділі 4.6.3 зазначено основні зміни у функціонуванні фонем, пов'язані з вживанням ірландської говірки. Заміна одних фонем іншими призвела до зміни розподілу фонем, що спричинило істотні розходження за більшістю груп фонем. Таблиці А.46 – А.48 Додатку А представляють дані для всіх трьох випадків.

Порівнюючи художню прозу Байрона з газетним стилем, істотні розходження отримано за трьома групами фонем для випадку позиції фонемі на початку слова. Це групи середньоязикових, задньоязикових і щілинних. Для випадків не врахування позиції фонемі в слові та позиції фонемі в кінці слова художня проза і

газетний стиль істотно відрізняються за чотирма групами фонем. Це відповідно групи: губних, сонорних, щілинних і зімкнених; губних, задньоязикових, сонорних і щілинних [147].

Прозові зауваги Байрона, як вже зазначалося раніше, торкаються суспільно-політичної сфери діяльності, і тому лексичні одиниці, які використовуються у газетному стилі, потрапили у досліджувану художню прозу. Про це свідчать показники середньої частоти груп приголосних фонем, за якими художня проза Байрона відрізняється від газетного стилю за порівняно незначною кількістю груп фонем. Таблиці А.49 – А.51 Додатку А представляють дані для всіх трьох випадків.

Отримані результати статистичного аналізу доцільно подати у вигляді структурної схеми вдосконаленої статистичної моделі підстилевої диференціації текстів (рис.4.13), яка репрезентує закономірності співвідношення якісних та кількісних характеристик англійських приголосних фонем у межах фонологічної підсистеми системи художнього стилю.

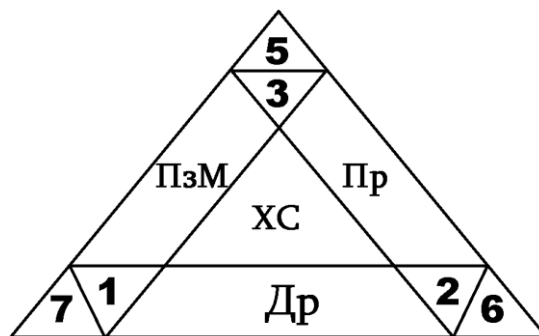


Рисунок 4.13 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі підстилевої диференціації художнього стилю на фонологічному рівні

На рис. 4.13 використано такі позначення: ПзМ – поезія Т. Мура, Пр – художня проза Дж. Г. Байрона, Др – драматургія Б. Шоу, ХС – художній стиль; 1, 2, 3 – невелика кількість груп фонем, за якими встановлено істотні відмінності (тексти схожі); 5, 6, 7 – велика кількість груп фонем, за якими встановлено істотні відмінності (тексти відмінні).

Дана статистична модель з більшою достовірністю робить можливим встановлення приналежності досліджуваного тексту до певного підстилю одного стилю за кількістю груп приголосних фонем, за якими встановлено істотні відмінності [148].

4.6.3 Результати дослідження фоностатистичної структури драми

Матеріалом дослідження підстилю драми є п'єса Бернарда Шоу "John Bull's Other Island" (1904) [149]. Поява цього твору викликана кельтським відродженням, метою якого було сприяння розквіту національного ірландського мистецтва – музики, літератури, живопису. П'єса є патріотичною. Драматург вдало відтворив національний колорит Ірландії через пейзажні описи ірландської природи та образи ірландців, які говорять ірландською говіркою.

За результатами аналізу лексичного рівня мови твору виявлено значну кількість реплік, які відображають ірландську говірку. Ці лексичні відмінності можуть, до певної міри, впливати на частоту вживання фонем. Так, наприклад, фонема *f* не вживається у прийменнику *of*; фонема *n* вживається у словах із закінченням *ing*; фонема *m* вживається замість фонем *v* та *t* у словах *give* і *let*; дифтонг *ai* замінено фонемою *i*: у займеннику *my*; *i* вживається замість *e* у словах *when*, *respect*, *expenses* та багатьох інших словах; слова *dhrink*, *dhróp*, *wondher*, *sthrong*, *counthry* та інші слова вживаються замість *drink*, *drop*, *wonder*, *strong* і *country* [150]. Для об'єктивної оцінки специфіки фонологічного рівня мови твору, доцільно представити результати статистичного обстеження даної вибірки.

Всебічне висвітлення порівняння підстилю драми з поезією Байрона і Мура, художньою прозою Байрона та з розмовним стилем є представлене у розділах 4.6; 4.6.1; 4.6.2. Доречно навести основні результати порівняння підстилю драми з кожним із вищезазначених стилів та підстилів з метою цілісної характеристики підстилю. Підстиль драми істотно відрізняється від поезії Байрона за шістьма групами фонем для випадків не врахування позиції фонем в слові та позиції фонем на початку слова, а для випадку позиції фонем в кінці слова – за

п'ятьма. У даному випадку істотна відмінність за більшістю груп фонем пояснюється більшою стійкістю границь підстилю драми до проникнення елементів суспільно-політичної лексики та реплік з ірландською говіркою (Таблиці А.22 – А.24 Додатку А).

Підстиль драми істотно відрізняється від поезії Мура за сімома групами фонем для випадків не врахування позиції фонемі в слові та позиції фонемі на початку слова, а для випадку позиції фонемі в кінці слова – лише за двома. Відмінність підстилю драми від поезії Мура для перших двох випадків можна пояснити вищезгаданими причинами його відмінності від поезії Байрона. Результати отримані для третього випадку свідчать, у більшій мірі, про закономірності функціонування фонем даної фонологічної системи, ніж про подібність підстилів в даній позиції (Таблиці А.34 – А.36 Додатку А).

При порівнянні підстилів драми і художньої прози Байрона істотні розходження виявлено за шістьма групами приголосних фонем для випадків позиції фонемі на початку слова і позиції фонемі в кінці слова та за сімома групами у випадку не врахування позиції фонемі в слові. Різна тематична спрямованість порівнюваних творів, яка зумовила відмінність лексичного складу, та вплив ірландської говірки пояснюють істотну відмінність за більшістю груп фонем для всіх трьох випадків (Таблиці А.46 – А.48 Додатку А).

Однакові результати отримано для всіх трьох випадків при порівнянні підстилю драми з розмовним стилем. Вибірки істотно відрізняються за чотирма групами приголосних фонем. Деяка близькість порівнюваних вибірок є очевидною, бо елементи розмовного стилю, у значній мірі, проникають у драму Шоу [151]. Незважаючи на вплив ірландської говірки, зазначена близькість підстилю і стилю підтверджується числовими показниками, отриманими у результаті проведеного статистичного обстеження (Таблиці А.10 – А.12 Додатку А).

Порівняння підстилю драми з газетним та науковим стилями не розглядається у попередніх підрозділах і тому отримує докладне висвітлення у даному підрозділі. Для випадку не врахування позиції фонемі в слові підстиль драми і газетний стиль істотно відрізняються за шістьма групами фонем. Це групи губних,

передньоязикових, середньоязикових, носових, сонорних і щілинних. Для випадку позиції фонему в кінці слова істотні розходження виявлено за п'ятьма групами фонем. Це групи губних, передньоязикових, сонорних, щілинних і зімкнених. Для випадку позиції фонему на початку слова вибірки істотно відрізняються за чотирма групами фонем. Це групи середньоязикових, задньоязикових, носових і сонорних [152]. Слід зазначити, що зменшення кількості груп фонем, за якими істотно розрізняються порівнювані підстиль і стиль, зумовлене проникненням суспільно-політичної лексики в драму Шоу. Елементи ірландської говірки розмовного стилю в драмі Шоу, з одного боку, та значно більша кількість слів із суспільно-політичної сфери вжитих у газетному стилі, з іншого – це ті фактори, які спричинили вищезгадані відмінності вибірок. Числові дані представлено у таблицях А.55 – А.57 Додатку А.

При зіставленні підстилю драми Шоу з науковим стилем істотні розходження виявлено за сімома групами фонем для випадку позиції фонему на початку слова. Це групи губних, передньоязикових, середньоязикових, задньоязикових, носових, сонорних і зімкнених. Для випадку не врахування позиції фонему в слові вибірки істотно відрізняються за шістьма групами фонем. Це групи передньоязикових, середньоязикових, задньоязикових, щілинних, зімкнених і сонорних. Для випадку позиції фонему в кінці слова істотні розходження встановлено за шістьма групами фонем. Це групи губних, передньоязикових, задньоязикових, сонорних, щілинних і зімкнених.

Розмежування даних стилів за більшістю груп фонем зумовлене специфікою кожного із порівнюваних текстів. Значна кількість термінів з галузі технічної фізики протистоїть сукупності лексичних одиниць, яка охоплює елементи, які вживаються у художній прозі, суспільно-політичній сфері, і елементи розмовної мови, деякі з яких є лексичними одиницями ірландської говірки. Числові дані дослідження представлено у таблицях А.58 – А.60 Додатку А.

Результати диференціації підстилю драми, художнього стилю і розмовного стилю, отримані з довірчою ймовірністю 0,95, репрезентує вдосконалена статистична модель стилєвої диференціації текстів (рис. 4.13).

Дана статистична модель дає змогу з більшою достовірністю здійснити стилеву атрибуцію досліджуваного тексту за кількістю груп приголосних фонем, за якими виявлено істотні відмінності [153].

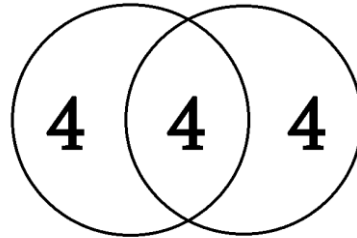


Рис. 4.13 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі стилєвої диференціації текстів підстилю драми художнього стилю і розмовного стилю на фонологічному рівні

На рис. 4.13 використано такі позначення: 4 – істотні та неістотні розбіжності встановлено за чотирма групами приголосних фонем.

4.7 Результати дослідження фоностатистичної структури газетного стилю

Матеріалом дослідження є газетні повідомлення, заголовки і оголошення, які складають основу газетного стилю. Рональд Картер і Вольтер Неш висловлюють думку, що стиль газетних повідомлень повинен формуватись із елементів так званого ‘серцевинного’ лексичного рівня. До ‘серцевинного’ лексичного рівня дослідники відносять немарковані елементи мови, тобто ті мовні елементи, які складають основний словниковий запас носія мови і відзначаються найвищим рівнем частоти вживання [154]. Важко провести чітку лінію розмежування між тими словами, які входять в ‘серцевинний’ пласт, і тими, які в нього не входять. Точніше можна говорити про те, що слово відноситься до ‘серцевинного’ рівня у більшій чи у меншій мірі.

Віднесення частини мовних одиниць до сфери немаркованості допомагає виокремити ті марковані елементи, які несуть додатковий відтінок значення. Саме

цей додатковий елемент у значенні слова дає змогу прагматично впливати на читача.

Вищезазначена специфіка дозволяє ствердити, що в цілому, для стилю газети є характерне поєднання немаркованої і маркованої лексики у різних пропорціях.

Системі функціональних стилів властиві взаємопроникнення і взаємозбагачення у процесі її постійного розвитку. Незважаючи на наявність елементів інших стилів у газетному стилі, він, як і решта стилів, характеризується набором суттєвих, диференційних ознак, які виокремлюють його в окремий функціональний стиль. Статистичні характеристики визначають специфіку розподілу елементів газетного стилю.

З метою визначення статистичних характеристик властивих стилю газети, досліджуються тексти (2018) (рубрика “Новини”) з англійської газети [155, 156].

У даному розділі всебічний аналіз отримує лише порівняння газетного стилю з науковим стилем, бо порівняння з іншими стилями представлено у підрозділах 4.5. і 4.6. З метою послідовного викладу матеріалу та розгляду специфіки газетного стилю на фонологічному рівні у систематизованому вигляді наведено короткий виклад вже проаналізованих у попередніх розділах даних, при цьому, посилаючись на відповідний розділ та таблиці.

Найбільш вдалим для встановлення істотних розходжень є випадок позиції фонем на початку слова при порівнянні газетного стилю з поезією Байрона. У цьому випадку вибірки істотно відрізняються за сімома групами фонем. Для випадку позиції фонем в кінці слова істотні розходження встановлено для чотирьох груп фонем. При не врахуванні позиції фонем в слові отримано істотні відмінності лише за однією групою фонем – передньоязикових, що свідчить про певні особливості функціонування фонем, які не залежать від позиції фонем в слові та про деяку близькість поезії Байрона і газетного стилю зумовлену спільними загальноживаними у щоденному спілкуванні словами (Таблиці А.25 – А.27 Додатку А).

Газетний стиль і поезія Мура відрізняються за більшістю груп приголосних фонем для випадків не врахування позиції фонем в слові та позиції фонем на

початку слова. Відповідно до результатів – це шість та сім груп фонем. Істотні розходження виявлено за чотирма групами фонем для випадку позиції фонем в кінці слова (Таблиці А.37 – А.39 Додатку А).

Виразна тенденція до зближення простежується при порівнянні художньої прози Байрона з газетним стилем [157]. У всіх трьох випадках істотні розходження виявлено за чотирма-трьома групами фонем. Ця подібність, очевидно, зумовлена наявністю спільних елементів в прозових заувагах Байрона поеми "Паломництво Чайльд Гарольда" і в досліджуваному газетному тексті (Таблиці А.49 – А.51 Додатку А).

Газетний стиль відрізняється від драми Б. Шоу в більшій мірі, ніж від художньої прози Байрона. Істотні розходження виявлено за шістьма групами фонем для випадку не врахування позиції фонем в слові, за п'ятьма групами фонем для випадку позиції фонем в кінці слова та за чотирма групами фонем для випадку позиції фонем на початку слова. Зазначені розходження у драмі Шоу спричинено діалектизмами (Таблиці А.55 – А.57 Додатку А).

У порівнянні з розмовним стилем газетний стиль істотно відрізняється від розмовного за шістьма групами фонем для випадків не врахування позиції фонем в слові та позиції фонем на початку слова [158]. Для випадку позиції фонем в кінці слова стилі істотно відрізняються за трьома групами фонем. Перші два випадки підтверджують вплив спеціальної газетної лексики, яка не вживається у щоденному спілкуванні на побутові теми (Таблиці А.13 – А.15 Додатку А).

У даному підрозділі зіставлення газетного і наукового стилів отримує всебічний розгляд, бо воно не представлено у попередніх підрозділах. Відмінність порівнюваних стилів, яка, природно, впливає з відмінності складу лексичних одиниць, що вживаються в двох різних сферах людської діяльності, отримала чіткіше підтвердження для випадку позиції фонем на початку слова, в якому стилі істотно відрізняються за шістьма групами фонем: губних, передньоязикових, середньоязикових, задньоязикових, носових і щілинних. Для випадку позиції фонем в кінці слова істотні розходження виявлено для передньоязикових, задньоязикових, носових і зімкнених. Стилi розрізняються ще за меншою кількістю

груп фонем для випадку не врахування позиції фонем в слові. Це групи передньоязикових, носових і зімкнених [159].

Вищерозглянуті результати, отримані з довірчою ймовірністю 0,95, підтверджують взаємопроникнення стилів та схожість фонематичного наповнення у кожній позиції зокрема. Таблиці А.61 – А.63 Додатку А представляють дані досліджень для трьох випадків.

Для підвищення достовірності та автоматизації стилевої, підстилевої та авторської атрибуції тексту реалізовано вдосконалену статистичну модель визначення стилерозрізняльної здатності груп губних, передньоязикових, середньоязикових і задньоязикових фонем за методом гіпотез при диференціації попарно зіставлених текстів поезії Байрона, Мура, художньої прози Байрона, драми Шоу, розмовного стилю, газетного стилю і наукового стилю у двох наступних випадках позиції фонем в слові: I – не враховуючи позицію фонем в слові; II – фонема на початку слова. Показники стилерозрізняльної здатності груп фонем наведено у таблицях 4.4, 4.5, 4.6, 4.7.

Таблиця 4.4

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	-	-	-	-	+	+	+	+	-	+	+	+		

Таблиця 4.5

Результати визначення стилерозрізняльної здатності групи передньоязикових 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	+	+	+	+	+	-	+	-	+	+	+	+		

Таблиця 4.6

Результати визначення стилерозрізняльної здатності групи середньоязикових 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	+	+	+	+	+	+	-	+	+	+	+	+		

Таблиця 4.7

Результати визначення стилерозрізняльної здатності групи задньоязикових 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	+	+	+	-	+	-	-	-	+	+	+	+		

У таблиці 4.7 використано такі позначення: I – не враховуючи позицію фонем в слові; II – фонема на початку слова; позначення текстів: ПзБ – поезія Дж. Г. Байрона, ПзМ – поезія Т. Мура, Пр – художня проза Дж. Г. Байрона, Др – драматургія Б. Шоу, РС – розмовний стиль, ГС – газетний стиль, НС – науковий стиль.

Істотні відмінності між зіставленими стилями за групою приголосних фонем позначено знаком «+», неістотні – знаком «-».

Стилерозрізняльна здатність груп передньоязикових та задньоязикових груп фонем наведена у вигляді структурної схеми представлення результатів вдосконаленої статистичної моделі (рис. 4.14)

ПзБ	РС	ПзБ РС	
ПзМ	РС	ПзМ РС	
Пр РС		Пр РС	
Др РС		Др РС	
РС	ГС	РС	ГС
РС	НС	РС	НС
ПзБ	Пр	ПзБ	Пр
ПзБ	Др	ПзБ Др	
ПзБ	ГС	ПзБ ГС	
ПзБ	НС	ПзБ НС	
ПзМ	Пр	ПзМ Пр	
ПзМ	Др	ПзМ Др	
ПзМ	ГС	ПзМ ГС	
ПзМ	НС	ПзМ НС	
ПзБ ПзМ		ПзБ	ПзМ
Пр Др		Пр Др	
Пр ГС		Пр	ГС
Пр	НС	Пр	НС
Др ГС		Др	ГС
Др	НС	Др	НС

Рисунок 4.14 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі визначення стилерозрізняльної здатності передньоязикових та задньоязикових груп фонем

На рис. 4.14 використано такі позначення: ПзБ – поезія Дж. Г. Байрона, ПзМ – поезія Т. Мура, Пр – художня проза Дж. Г. Байрона, Др – драматургія Б. Шоу, РС – розмовний стиль, ГС – газетний стиль, НС – науковий стиль.

Відсутність проміжку – неістотні відмінності між досліджуваними текстами; наявність проміжку – істотні відмінності.

Вдосконалена статистична модель дає змогу з довірчою ймовірністю 0,95 визначити стилеву атрибуцію досліджуваного тексту [160].

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 трьох випадків позиції фонем в слові, що зображено на рис. 4.15. Результати реалізації вдосконаленої статистичної моделі стилеві

диференціації текстів свідчать про середні істотні відмінності газетного стилю від розмовного.

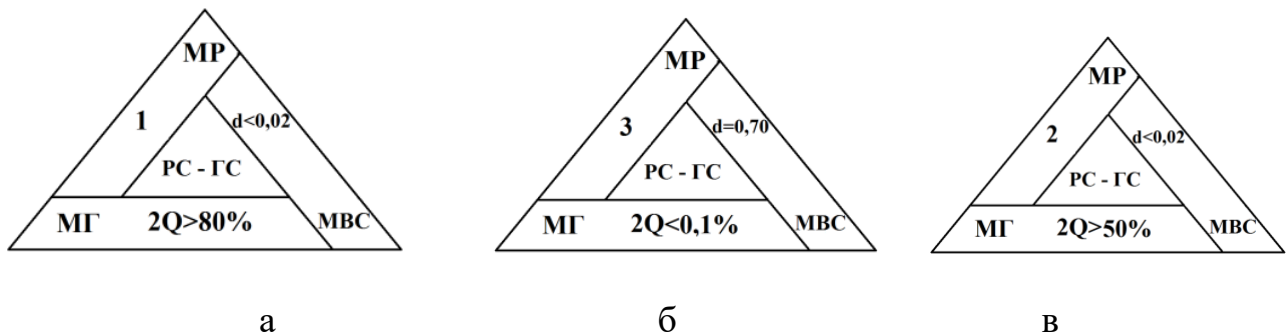


Рисунок 4.15 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі стилєвої диференціації текстів за методом гіпотез (МГ), методом ранжування (МР) та методом визначення відстаней між стилями (МВС) для зіставлених розмовного і газетного стилів за групою сонорних фонем для трьох 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

У даному зіставленні стилів кожен із застосованих трьох методів забезпечив отримання середніх істотних відмінностей (числові показники зображено на рис. 4.15), що підтверджує дієвість кожного методу зокрема, а їх поєднання дозволяє підвищити достовірність стилєвої атрибуції тексту.

4.8 Результати дослідження фоностатистичної структури наукового стилю

Точність викладу думки є однією з основних ознак наукового стилю. Слова вживаються або в основних, або в термінологічних значеннях, але не в експресивно-образних. Фактори експресивності і авторської індивідуальності є основними при визначенні відмінностей між науковим стилем, з одного боку, і поезією, художньою прозою та драмою, з другого боку. Слід зауважити, що для поезії, художньої прози та драми, своєрідність авторського слова є виявом авторської майстерності.

У науковому стилі, індивідуальний стиль автора відходить на задній план, зазнаючи впливу традиційних і стереотипних форм наукового викладу.

Факторами, за якими розмежують науковий і художній стилі, є об'єктивна дійсність і відправник мовлення. Очевидно, що фактор об'єктивної дійсності є вирішальним для наукового стилю, хоча, інколи, необхідно враховувати і фактор відправника мовлення. Це стосується тих випадків, коли автор володіє письменницькими здібностями і використовує нетрадиційну форму викладу своїх суджень та інтерпретації результатів дослідження. Проте такі випадки трапляються рідко і тому дія фактору відправника мовлення є, в цілому, незначною [161].

Складність і суперечливість наукового стилю полягає у тому, що він об'єднує наукові тексти, які відносяться до різних сфер дослідження. Відсутність однієї тематики порушує проблему монолітності наукового стилю. Ця проблема отримала різне трактування багатьма лінгвістами. Зокрема, Д. Кристал і Д. Деві вважають, що система мовних засобів загалом збігається у різних тематичних текстах, якщо не брати до уваги термінологічну лексику [162]. Проблема визначення найхарактерніших випадків порушення стилістичних норм у науковому стилі докладно висвітлена у дослідженні О. П. Левченко [163]. Наукові статті з різних галузей фізики (*Acta Physica Polonica* 2018) [164] вибрано для характеристики специфіки наукового стилю.

Докладний розгляд порівняння наукового стилю з іншими стилями представлено у підрозділах 4.5, 4.6, 4.7. З метою узагальнення та систематизації опису отриманих результатів, наводиться короткий виклад порівняння наукового стилю з поезією, художньою прозою, драмою, газетним та розмовним стилями.

Істотні розходження виявлено за всіма групами фонем у порівнянні з поезією Байрона для випадку позиції фонем на початку слова. Майже тотожний результат отримано для випадку позиції фонем в кінці слова, в якому лише за групою сонорних не виявлено істотних розходжень. Не враховуючи позицію фонем в слові, у текстових зіставленнях встановлено істотні розходження лише для груп губних і зімкнених (Таблиці А.28 – А.30 Додатку А).

Порівнюючи науковий стиль з поезією Мура, отримано такі результати: істотні розходження встановлено за шістьма групами фонем для випадків не врахування позиції фонем в слові та позиції фонем на початку слова. Вибірки істотно відрізняються за п'ятьма групами фонем для випадку позиції фонем в кінці слова. Отже, загалом науковий стиль істотно відрізняється від поезії за більшістю груп фонем (Таблиці А.40 – А.42 Додатку А).

Науковий стиль і художня проза Байрона істотно відрізняються за шістьма групами фонем для всіх розглянутих випадків (Таблиці А.52 – А.54 Додатку А).

Порівняння наукового стилю і драми Шоу теж дало істотні розходження для більшості випадків. Для випадку позиції фонем на початку слова вибірки істотно відрізняються за сімома групами фонем. Для випадків не врахування позиції фонем в слові та для позиції фонем в кінці слова істотні розходження отримано за шістьма групами фонем (Таблиці А.58 – А.60 Додатку А).

Слід зазначити, що чітка відмінність наукового стилю простежується не лише у порівнянні його з підстилями художнього стилю, але й з розмовним стилем, матеріалом дослідження якого є щоденні, побутові тексти. Так, для випадків не врахування позиції фонем в слові та позиції фонем на початку слова стилі істотно відрізняються за сімома групами фонем. Для випадку позиції фонем в кінці слова істотні розходження встановлено за чотирма групами фонем (Таблиці А.16 – А.18 Додатку А).

У порівнянні з газетним стилем істотні відмінності встановлено за шістьма групами фонем лише для випадку позиції фонем на початку слова. Для випадків позиції фонем в кінці слова та її не врахування істотні відмінності отримано відповідно за чотирма та трьома групами фонем (Таблиці А.61 – А.63 Додатку А).

У попарному зіставленні художнього, розмовного, газетного 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) і розмовного стилів (розмовно-побутова лексика) за групою передньоязикових фонем, використовуючи розроблений метод комплексного аналізу диференціації фоностатистичних структур

стилів англійської мови (Таблиця 4.8). Одержані результати зумовлені насиченістю термінами технічної фізики текстів наукового стилю і відсутністю цих лексичних одиниць у текстах розмовного стилю. Істотні відмінності між стилями встановлено з довірчою ймовірністю 0,95 [165]. Результати одержано запропонованим поєнанням статистичних методів та використанням потужного непараметричного критерію Колмогорова-Смірнова.

Таблиця 4.8

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

Вищесказане дає змогу ствердити, що науковий стиль істотно відрізняється за більшістю груп фонем від інших функціональних стилів не для всіх трьох випадків позиції фонем в слові. Зазначена відмінність зумовлена різною стилерозрізняльною здатністю груп приголосних фонем у різних позиціях фонем у слові.

Результати попарного порівняння досліджуваних стилів представлено у таблиці А.64 Додатку А.

Відмінність наукового стилю від розмовного, художнього і газетного встановлено за стилерозрізняльною здатністю групи губних фонем (Таблиці 4.9 – 4.10; Рис. 4.15).

Результати попарних зіставлень досліджуваних текстів для випадку довільної позиції фонем в слові наведено у таблицях 4.9 і 4.10.

У таблицях використано такі позначення текстів: ПЗБ – поезія Дж. Г. Байрона, ПЗМ – поезія Т. Мура, Пр – художня проза Дж. Г. Байрона, Др – драматургія Б. Шоу, РС – розмовний стиль, ГС – газетний стиль, НС – науковий стиль.

Результати попарних порівнянь досліджуваних текстів для випадку позиції фонем на початку слова представлено у таблицях 4.11 і 4.12.

Таблиця 4.9

Істотні відмінності досліджуваних текстів за групою губних фонем для випадку довільної позиції фонем в слові.

Діючий фактор	Зіставлені тексти
Фактор авторської манери викладу	–
Фактор підстилю	ПзМ-Др; Пр-Др; ПзБ-Пр; ПзБ-Др
Фактор стилю	РС-Пр; РС-Др; РС-НС; ПзБ-НС; ПзМ-ГС Др-ГС; ПзМ-НС; Пр-ГС; Пр-НС

Таблиця 4.10

Випадкові відмінності досліджуваних текстів за групою губних фонем для випадку довільної позиції фонем в слові.

Діючий фактор	Зіставлені тексти
Фактор мови	РС-ПзБ; РС-ПзМ; РС-ГС; ПзБ-ГС; Др-НС; ГС-НС
Фактор підстилю	ПзБ- ПзМ
Фактор стилю	Пр- ПзМ

Таблиця 4.11

Істотні відмінності досліджуваних текстів за групою губних фонем для випадку позиції фонем на початку слова.

Діючий фактор	Зіставлені тексти
Фактор авторської манери викладу	–
Фактор підстилю	ПзБ-Пр; Пр-ПзМ; ПзМ-Др; Пр-Др; ПзБ-Др
Фактор стилю	РС-Пр; РС-Др; РС-НС; РС-ГС; ПзБ-ГС; ПзБ-НС; ПзМ-ГС; Пр-НС; Др-НС; ПзМ-НС; ГС-НС

Таблиця 4.12

Випадкові відмінності досліджуваних текстів за групою губних фонем для випадку позиції фонем на початку слова.

Діючий фактор	Зіставлені тексти
Фактор мови	РС-ПзБ; РС-ПзМ; Пр-ГС; Др-ГС
Фактор підстилю	ПзБ- ПзМ
Фактор стилю	–

Результати попарних зіставлень досліджуваних стилів для випадку позиції фонем в кінці слова представлено у таблицях 4.13 і 4.14.

Відмінність та схожість лексико-семантичного наповнення зіставляваних текстів наукового стилю з текстами розмовного, художнього і газетного стилів визначено реалізацією вдосконаленої статистичної моделі стилевої, підстилевої, та авторської диференціації текстів за методом гіпотез (критерієм Стьюдента), яка дала змогу встановити кількість груп приголосних фонем, за якими стилі відрізняються істотно. Розмежування стилів за кількістю груп приголосних фонем здійснено вперше для підвищення достовірності диференціації стилів. У таблиці 4.15 наведено результати для випадку не врахування позиції фонем в слові.

Таблиця 4.13

Істотні відмінності досліджуваних текстів за групою губних фонем для випадку позиції фонем в кінці слова.

Діючий фактор	Зіставлені тексти
Фактор авторської манери викладу	ПзБ-ПзМ
Фактор підстилю	ПзБ-Пр; ПзБ-Др
Фактор стилю	РС-Пр; РС-НС; ПзБ-НС; ПзМ-ГС Др-ГС; ПзМ-НС; Пр-ГС; Пр-НС; Др-НС

Таблиця 4.14

Випадкові відмінності досліджуваних текстів за групою губних фонем для випадку позиції фонем в кінці слова.

Діючий фактор	Зіставлені тексти
Фактор мови	РС-ПзБ; РС-ПзМ; РС-ГС; ПзБ-ГС; ГС-НС РС-Др
Фактор підстилю	–
Фактор стилю	Пр- ПзМ; ПзМ-Др;Пр-Др

Значні істотні відмінності встановлено за 6 – 7 групами фонем; середні за 4 – 5 групами фонем; незначні за 1 – 3 групами фонем. Одержані результати дають змогу зробити висновок про потужність критерію Стьюдента. У зіставленні поезії Байрона з науковим стилем, істотні відмінності встановлено лише за двома групами фонем, а з газетним стилем – за одною. Таким чином, критерій Стьюдента є менш потужним, ніж критерії Колмогорова-Смірнова та хі-квадрат.

Таблиця 4.15

Позиція фонем в слові неврахована

Стиль	ПзБ	ПзМ	Пр	Др	ГС	НС	РС
	І	І	І	І	І	І	І
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	-

У таблиці 4.15 використано такі позначення: ПзБ – поезія Дж. Г. Байрона, ПзМ – поезія Т. Мура, Пр – художня проза Дж. Г. Байрона, Др – драматургія Б.Шоу, РС – розмовний стиль, ГС – газетний стиль, НС – науковий стиль.

Реалізована вдосконалена статистична модель визначення стилерозрізняльної здатності групи губних фонем для трьох випадків позиції фонему в слові за методом гіпотез при розмежуванні попарно зіставлених текстів розмовного, художнього, газетного і наукового стилів дає змогу наголосити на переважанні істотних відмінностей, що підтверджує високу стилерозрізняльну здатність групи губних фонем (рис. 4.16).

Врахування усіх випадків позиції фонему в слові підвищує достовірність визначення стилерозрізняльної здатності групи губних фонем і дає змогу підвищити рівень автоматизації авторської атрибуції за цією групою фонем.

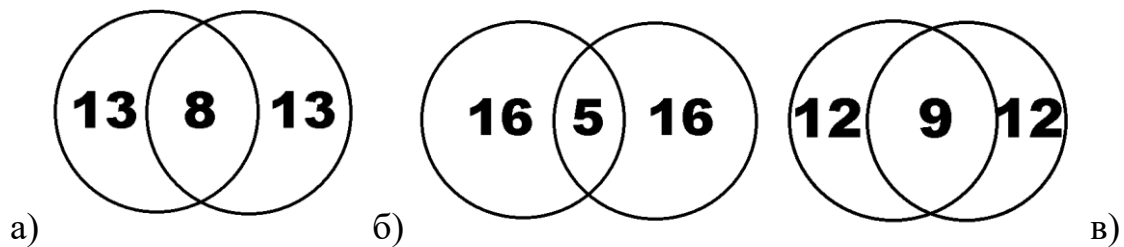


Рисунок 4.16 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі визначення стилерозрізняльної здатності групи губних фонем у трьох випадках позиції фонему в слові

На рис. 4.16 використано такі позначення: а) довільна позиція фонему в слові; б) фонему на початку слова; в) фонему в кінці слова. Істотні відмінності – у 13(а); 16(б) і 12(в) парах зіставлених текстів. Випадкові відмінності – у 8(а); 9(б) і 6(в) парах зіставлених текстів.

Результати реалізації вдосконаленої статистичної моделі визначення стилерозрізняльної здатності всіх восьми груп приголосних фонем за критеріями Стюдента, Колмогорова-Смірнова та хі-квадрат з довірчою ймовірністю 0,95 наведено у таблиці 4.16.

Застосування трьох критеріїв диференціації авторських стилів, одним з яких є потужний критерій (критерій Колмогорова-Смірнова), дало змогу підвищити достовірність авторської атрибуції тексту. За результатами наведеними у таблиці 4.16 групи передньоязикових, задньоязикових і носових фонем мають найвищу

стилерозрізняльну здатність і за цими групами можна підвищити рівень автоматизації атрибуції тексту.

Таблиця 4.16

Результати визначення стилерозрізняльної здатності восьми груп приголосних фонем

Групи фонем	Критерій Стьюдента	Критерій Колмогорова-Смірнова	Критерій хі-квадрат
Губні	-	+	+
Передньоязикові	+	+	+
Середньоязикові	-	+	-
Задньоязикові	+	+	-
Носові	+	+	+
Сонорні	-	+	+
Щілинні	-	+	+
Зімкнені	-	+	+

Таким чином, вдосконалені статистичні моделі дають уможливають підвищення достовірності та рівню автоматизації стилевої, підстилевої та авторської атрибуції тексту.

4.9 Результати дослідження фоностатистичної структури авторського стилю

Однією з важливих задач математичної лінгвістики є встановлення авторства тексту. Встановлення авторства тексту на фонологічному рівні можна здійснювати за такими критеріями диференціації: абсолютна частота груп фонем; відносна частота груп фонем; середня частота груп фонем [166, 167]. Тексти можна диференціювати за: однією групою фонем; двома-трьома групами фонем; всіма, виокремленими за певною класифікацією, групами фонем.

Найефективнішим підходом до розв'язання вищезазначеної задачі є розмежування текстів за однією-трьома групами фонем, за якими встановлено істотні відмінності у більшості випадків зіставлення текстів різних авторів [168]. Таке спрощення процесу авторської атрибуції тексту є зменшенням кількості груп фонем, за якими встановлено істотні відмінності між текстами різних авторів.

На частоту вживання фонем впливають мовні, стилеві і авторські чинники. При незначній маркованості тексту стилевими та авторськими елементами найвищу авторорозрізняльну здатність має та група фонем, яка характеризується високою частотою вживання у мові загалом [169, 170].

Завданням дослідження є виявлення істотних відмінностей у випадку приналежності текстів до одного стилю (художній) та підстилю (поезія), але відрізняються авторським стилем. У цьому випадку 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 (Таблиця 4.17).

Істотна відмінність авторських стилів Байрона і Мура встановлена за різницею значень загальної стилевої маркованості зіставлених текстів поезії і дорівнює 1,8 (Таблиця 4.18).

Отриманий результат підтверджує відмінність індивідуальних для кожного автора особливостей манери викладу у текстах схожої тематики у межах одного літературного напрямку – романтизму [171].

Таблиця 4.17

Результати визначення значення загальної стилевої маркованості за методом гіпотез

Зіставлення з іншим стилем	Зіставлення з іншим підстилем	Зіставлення з іншим автором	Значення загальної стилевої маркованості
ПзБ-Рс – 16	ПзБ-Пр – 17	ПзБ-ПзМ – 7	ПзБ – 13,3
ПзМ-Рс – 15	ПзМ-Пр – 13	ПзМ-ПзБ – 7	ПзМ – 11,5

У таблиці 4.17 використано такі позначення: ПзБ – поезія Дж. Г. Байрона, ПзМ – поезія Т. Мура, Пр – художня проза Дж. Г. Байрона, РС – розмовний стиль.

Встановлений показник відмінності авторських стилів Байрона і Мура свідчить про дієвість багатофакторного методу, який дає змогу підвищити достовірність ідентифікації автора поетичного твору встановленням ступеню дії трьох факторів (стилю, підстилю та авторської манери викладу).

Встановлення диференційних характеристик авторського стилю здійснено на матеріалі двох текстів з художнього стилю: “Грозовий перевал” Е. Бронте (“Wuthering Heights” by E. Bronte) [172] та “Ярмарок марнославства” В. Текерея (“Vanity Fair” by W. Thackeray) [173]. Результати, отримані за критерієм Стюдента, зображено на рис. 4.17. Сучасна класифікація англійських приголосних фонем включає: губні – labial; передньоязикові – dorsal (дорсальні); передньоязикові, апікальні, альвеолярні – coronal; задньоязикові – velar (велярні); носові – nasal; сонорні – sonorous; щілинні – fricative; зімкнені – stop [174, 175].

Таблиця 4.18

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

Суттєві відмінності встановлено за групами носових і сонорних фонем. Результати отримані за критерієм Колмогорова-Смірнова зображено на рис. 4.18. Суттєві відмінності встановлено за всіма восьмома групами фонем. Результати отримані за критерієм хі-квадрат зображено на рис. 4.19.

Результати представлені на рис. 4.18, рис. 4.19 і рис. 4.20 показують авторорозрізняльну здатність груп фонем у співставленні двох текстів з художнього стилю: “Грозовий перевал” Е. Бронте (“Wuthering Heights” by E. Bronte) та “Ярмарок марнославства” В. Текерея (“Vanity Fair” by W. Thackeray). Загальну авторорозрізняльну здатність груп фонем, отриману за трьома критеріями, представлено у таблиці 4.19.

```

24     "Statistic Criterion": "STUDENT",
25     "Results of Pearson test of sample1": {
26         "LABIAL": false,
27         "VELAR": false,
28         "FRICATIVE": true,
29         "NASAL": true,
30         "SONOROUS": true,
31         "CORONAL": false,
32         "DORSAL": false,
33         "STOP": false
34     },
35     "Results of Pearson test of sample2": {
36         "LABIAL": false,
37         "VELAR": true,
38         "FRICATIVE": true,
39         "NASAL": true,
40         "SONOROUS": true,
41         "CORONAL": false,
42         "DORSAL": false,
43         "STOP": true
44     },
45     "Statistic Results": {
46         "LABIAL": null,
47         "VELAR": null,
48         "FRICATIVE": true,
49         "NASAL": false,
50         "SONOROUS": false,
51         "CORONAL": null,
52         "DORSAL": null,
53         "STOP": null
54     },
55     "Error message": []
56 }

```

Рисунок 4.17 – Результати за критерієм Стьюдента

```

/,
"Statistic Criterion": "KOLMOGOROV_SMIRNOV",
"Statistic Results": {
    "LABIAL": false,
    "VELAR": false,
    "FRICATIVE": false,
    "NASAL": false,
    "SONOROUS": false,
    "CORONAL": false,
    "DORSAL": false,
    "STOP": false
},
"Error message": []

```

Рисунок 4.18 – Результати за критерієм Колмогорова-Смірнова


```

"Statistic Criterion": "CHI_SQUARE",
"Statistic Results": {
  "LABIAL": true,
  "VELAR": true,
  "FRICATIVE": false,
  "NASAL": true,
  "SONOROUS": true,
  "CORONAL": true,
  "DORSAL": false,
  "STOP": false
},
"Error message": []

```

Рисунок 4.19 – Результати за критерієм хі-квадрат

Суттєві відмінності встановлено за групами щілинних, дорсальних і зімкнених фонем.

Таблиця 4.19

Результати визначення авторо-розрізняльної здатності груп фонем

Група фонем	Критерій Стюдента	Критерій Колмогорова- Смірнова	Критерій хі- квадрат
велярні		+	
щілинні		+	+
носові	+	+	
сонорні	+	+	
передньоязикові		+	
дорсальні		+	+
зімкнені		+	+
губні		+	

У наступному зіставленні порівнюються тексти з публіцистичного стилю – статті С. Логан і Д. Вебстер з газети “Вільна Газета” (“Freedom Paper”, papers by S. Logan and D. Webster) [176, 177]. Результати отримані за критерієм Стюдента зображено на рис. 4.20.

Варто зазначити, що статті С. Логан і Д. Вебстер а також публічні промови Б. Обама і Д. Трампа (які розглядаються далі) [178, 179] репрезентують американський варіант англійської мови. Тексти британського і американського варіантів попарно не зіставляються.

```

23     },
24     "Statistic Criterion": "STUDENT",
25     "Results of Pearson test of sample1": {
26         "SONOROUS": true,
27         "FRICATIVE": false,
28         "CORONAL": false,
29         "DORSAL": true,
30         "STOP": false,
31         "LABIAL": false,
32         "NASAL": true,
33         "VELAR": true
34     },
35     "Results of Pearson test of sample2": {
36         "SONOROUS": false,
37         "FRICATIVE": true,
38         "CORONAL": false,
39         "DORSAL": true,
40         "STOP": false,
41         "LABIAL": false,
42         "NASAL": true,
43         "VELAR": true
44     },
45     "Statistic Results": {
46         "SONOROUS": null,
47         "FRICATIVE": null,
48         "CORONAL": null,
49         "DORSAL": false,
50         "STOP": null,
51         "LABIAL": null,
52         "NASAL": false,
53         "VELAR": false
54     },
55     "Error message": []
56 }

```

Рисунок 4.20 – Результати за критерієм Стьюдента

Суттєві відмінності встановлено за групами носових, дорсальних і велярних фонем.

Результати отримані за критерієм Колмогорова-Смірнова зображено на рис. 4.21.

```

25     },
24     "Statistic Criterion": "KOLMOGOROV_SMIRNOV",
25     "Statistic Results": {
26         "SONOROUS": false,
27         "FRICATIVE": false,
28         "CORONAL": false,
29         "DORSAL": false,
30         "STOP": false,
31         "LABIAL": false,
32         "NASAL": false,
33         "VELAR": false
34     },
35     "Error message": []
36 }

```

Рисунок 4.21 – Результати за критерієм Колмогорова-Смірнова

Суттєві відмінності встановлено за всіма восьмима групами фонем. Результати отримані за критерієм χ^2 зображено на рис. 4.22.

```

24     "Statistic Criterion": "CHI_SQUARE",
25     "Statistic Results": {
26         "SONOROUS": false,
27         "FRICATIVE": false,
28         "CORONAL": true,
29         "DORSAL": false,
30         "STOP": false,
31         "LABIAL": false,
32         "NASAL": false,
33         "VELAR": true
34     },
35     "Error message": []
36 }

```

Рисунок 4.22 – Результати за критерієм χ^2

Суттєві відмінності встановлено за групами сонорних, щілинних, дорсальних, зімкнених, губних і носових фонем.

Результати представлені на рис. 4.20, рис. 4.21 і рис. 4.22 показують авторо-розрізняльну здатність груп фонем у співставленні двох текстів з публіцистичного стилю – статей С. Логан і Д. Вебстера з газети “Вільна Газета” (“Freedom Paper”, papers by S. Logan and D. Webster). Загальну авторо-розрізняльну здатність груп фонем отриману за трьома критеріями представлено у таблиці 4.20.

Таблиця 4.20

Результати визначення авторо-розрізняльної здатності груп фонем

Група фонем	Критерій Стьюдента	Критерій Колмогорова-Смірнова	Критерій хі-квадрат
велярні	+	+	
щілинні		+	+
носові	+	+	+
сонорні		+	+
передньоязикові		+	
дорсальні	+	+	+
зімкнені		+	+
губні		+	+

Результати отримані за трьома критеріями представлено у таблиці 4.21. Результати підтверджують здатність груп фонем розрізняти тексти з публіцистичного стилю – публічних промов американського президента Д. Трампа і статей Д. Вебстер з газети “Вільна Газета” (Public Speeches by D. Trump and papers by D. Webster from “Freedom Paper”).

Таблиця 4.21

Результати визначення авторо-розрізняльної здатності груп фонем

Група фонем	Критерій Стьюдента	Критерій Колмогорова-Смірнова	Критерій хі-квадрат
велярні	+	+	+
щілинні	+	+	+
носові		+	+
сонорні		+	+
передньоязикові		+	
дорсальні		+	+
зімкнені		+	+
губні		+	+

Результати отримані за трьома критеріями представлено у таблиці 4.22. Авторо-розрізняльну здатність груп фонем встановлено для зіставлених текстів з

публіцистичного стилю – публічних промов американського президента Д. Трампа і статей С. Логан з газети “Вільна Газета” (Public Speeches by D. Trump and papers by S. Logan from “Freedom Paper”).

Таблиця 4.22

Результати визначення авторо-розрізняльної здатності груп фонем

Група фонем	Критерій Стьюдента	Критерій Колмогорова-Смірнова	Критерій хі-квадрат
велярні	+	+	+
щілинні		+	+
носові		+	
сонорні		+	+
передньоязикові		+	+
дорсальні		+	+
зімкнені	+	+	
губні		+	+

Зменшення кількості груп фонем, за якими встановлено істотні відмінності між текстами різних авторів, спрощує процес авторської атрибуції тексту [180, 181]. У зіставленні текстів представленому у таблицях 4.20, 4.21 і 4.22 авторську атрибуцію можна здійснити за групою велярних фонем.

Для розв’язання завдання підвищення достовірності авторської диференціації текстів використано вдосконалену статистичну модель авторської диференціації за трьома критеріями методу гіпотез – критеріями Стьюдента, Колмогорова-Смірнова та хі-квадрат [182 – 184]. За результатами тестування моделі встановлено, що найпотужнішим критерієм є критерій Колмогорова-Смірнова, за яким тексти відрізняються за всіма восьмима групами приголосних фонем. Результати реалізації моделі наведено у таблиці 4. 23.

Таблиця 4.23

Результати визначення авторської диференціації за методом гіпотез

Зіставлені авторські стилі	Критерій Стьюдента	Критерій Колмогорова-Смірнова	Критерій хі-квадрат
Обама-Трампа	-	8	6
Обама-Вебстер	-	8	6
Обама-Логан	-	8	6
Трампа-Вебстер	2	8	7
Трампа-Логан	2	8	6
Логан-Вебстер	3	8	6
Бронте-Текерей	2	8	3

Встановлені істотні відмінності між текстами Б. Обама, Д. Трампа, Д. Вебстера, С. Логан, Е. Бронте і В. Текерей за кількістю груп приголосних фонем за трьома критеріями з довірчою ймовірністю 0,95 підтверджують своєрідність специфіки авторської манери викладу у текстах схожої тематики і дають змогу підвищити достовірність диференціації текстів цих авторів.

Авторську атрибуцію тексту здійснено розробленим багатofакторним методом визначення ступеню дії авторської манери викладу за восьмима групами приголосних фонем у зіставленні текстів Б. Обама, Д. Трампа, Д. Вебстера і С. Логан (таблиця 4.24). Істотна відмінність авторських стилів текстів Б. Обама, Д. Трампа, Д. Вебстера і С. Логан встановлена за різницею значень загальної стилевої маркованості зіставлених текстів (Таблиця 4.25). Отриманий результат підтверджує специфіку індивідуальних для кожного автора особливостей манери викладу у текстах схожої тематики у межах публіцистичного стилю.

Встановлені відмінності авторських стилів Б. Обама, Д. Трампа, Д. Вебстера і С. Логан з довірчою ймовірністю 0,95 свідчать, що багатofакторний метод дає змогу підвищити достовірність ідентифікації автора, визначаючи істотні відмінності у випадках дії трьох факторів (стилю, підстилю та авторської манери викладу).

Таблиця 4.24

Результати визначення значення загальної стилевої маркованості за методом гіпотез

Зіставлення з 1-м автором	Зіставлення з 2-м автором	Зіставлення з 3-м автором	Значення загальної стилевої маркованості
Обама-Трампа–14	Обама-Вебстера–14	Обама-Логана–14	Обама – 14
Трампа- Обама–14	Трампа-Вебстера–17	Трампа-Логана–16	Трампа–15,6
Логана-Вебстера–17	Логана-Обама–14	Логана-Трампа–16	Логана–15,6
Вебстера-Логана–17	Вебстера-Обама–14	Вебстера-Трампа–17	Вебстера–16

Таблиця 4.25

Результати зіставлення текстів Б. Обама, Д. Трампа, Д. Вебстера і С. Логана

Зіставлені тексти різних авторів за 8 групами фонем	Загальна стилева маркованість	Загальна стилева маркованість	Значення істотної відмінності
Обама-Трампа	Обама – 14	Трампа – 15,6	1,6
Обама-Вебстера	Обама – 14	Вебстера – 16	2
Обама-Логана	Обама – 14	Логана – 15,6	1,6
Трампа-Вебстера	Трампа – 15,6	Вебстера – 16	0,4
Трампа -Логана	Трампа – 15,6	Логана – 15,6	0
Вебстера-Логана	Вебстера – 16	Логана – 15,6	0,4

Крім використання потужних критеріїв, велику роль у підвищенні достовірності стилевої диференціації текстів відіграє врахування позиції фонем в слові. Результати реалізації вдосконаленої статистичної моделі стилевої диференціації наукового стилю і художньої прози Байрона за методом гіпотез у трьох випадках позиції фонем в слові зображено у вигляді схеми на рис. 4.23.

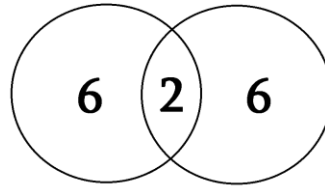


Рисунок 4.23 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі стилєвої диференціації текстів наукового стилю і художньої прози Байрона за методом гіпотез для 3 випадків позиції фонем в слові

На рис. 4.23 використано такі позначення: 2 – кількість груп фонем, за якими встановлено неістотні відмінності (тексти схожі); 6 – кількість груп фонем, за якими встановлено істотні відмінності (тексти відмінні).

Встановлення значних істотних відмінностей з довірчою ймовірністю 0.95 за 6 з 8 групами фонем є добрим результатом, який підтверджує відмінність текстів з різним лексико-семантичним наповненням і забезпечує високий рівень достовірності диференціації стилів.

Врахування позиції фонем в слові дає змогу підвищити достовірність диференціації текстів навіть при застосуванні менш потужного критерію (критерію Стьюдента). За результатами застосування вдосконаленої статистичної моделі авторської диференціації для текстів поезії Байрона і Мура за методом гіпотез для 3 випадків позиції фонем в слові отримано незначні істотні відмінності, що зображено на рис. 4.24.

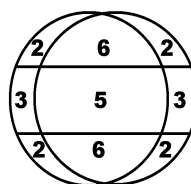


Рисунок 4.24 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі авторської диференціації текстів художнього стилю поезії Байрона і Мура

На рис. 4.24 використано такі позначення: 2, 3 – кількість груп фонем, за якими встановлено істотні відмінності (тексти відмінні); 5, 6 – кількість груп фонем, за якими встановлено неістотні відмінності (тексти схожі).

Встановлення істотних, хоч і незначних, (за 2 – 3 групами фонем) відмінностей між текстами тематично схожими є добрим результатом, який підтверджує дієвість вдосконаленої статистичної моделі та важливість врахування позиції фонем у слові для підвищення достовірності авторської атрибуції тексту.

Як і метод гіпотез, метод ранжування дав змогу отримати незначні істотні відмінності при диференціації текстів поезії Байрона і Мура для випадків позиції фонем на початку та в кінці слова. Реалізація вдосконаленої статистичної моделі авторської диференціації текстів за методом ранжування підтверджує схожість зіставлених текстів (рис. 4.25).

Одержання однакових результатів за двома різними методами (методом гіпотез і методом ранжування) підвищує достовірність диференціації авторських стилів, що підтверджує дієвість застосування поєднання методів. На основі результатів наведених у зазначених таблицях побудовано статистичну модель визначення авторо-розрізняльної здатності групи задньоязикових фонем (рис. 4.26).

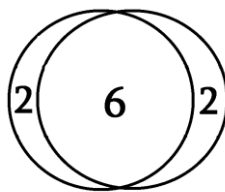


Рисунок 4.25 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі авторської диференціації текстів поезії Байрона і Мура за методом ранжування для випадків позиції фонем на початку та в кінці слова

На рис. 4.25 використано такі позначення: 2 – істотна різниця рангових показників (тексти відмінні); 6 – неістотна різниця рангових показників (тексти схожі).

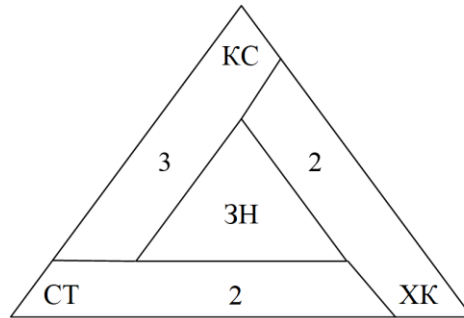


Рисунок 4.26 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі визначення авторо-розрізняльної здатності групи задньоязикових фонем

На рис. 4.26 використано такі позначення: КС – критерій Колмогорова-Смірнова, ХК – критерій хі-квадрат, СТ – критерій Стьюдента; ЗН – група задньоязикових фонем; 2, 2, 3 – кількість порівнянь, в яких група велярних фонем має авторо-розрізняльну здатність.

За результатами зіставлення текстів публіцистичного стилю – публічних промов американського президента Д. Трампа, статей С. Логан і Д. Вебстер з газети “Вільна Газета” побудовано статистичну модель визначення авторо-розрізняльної здатності групи щілинних фонем (рис. 4.27).

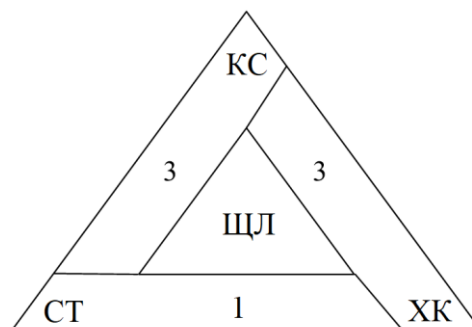


Рисунок 4.27 – Структурна схема представлення результатів реалізації вдосконаленої статистичної моделі визначення авторо-розрізняльної здатності групи щілинних фонем

На рис. 4.27 використано такі позначення: ЩЛ – група щілинних фонем; 1, 3, 3 – кількість порівнянь, у яких група щілинних фонем має авторо-розрізняльну здатність.

Розроблений метод комплексного аналізу диференціації фоностатистичних структур стилів, вдосконалена статистична модель визначення авторо-розрізняльної здатності групи приголосних фонем та програма на мові програмування Java забезпечили зменшення кількості груп фонем, за якими встановлено істотні відмінності між текстами різних авторів, що підвищує достовірність результатів і рівень автоматизації авторської атрибуції тексту.

4.10 Висновки до розділу 4

1. Розроблено структуру системи диференціації фоностатистичних структур стилів англійської мови, яка містить такі модулі: модуль введення/виведення даних; модуль перетворення англомовного тексту у транскрипційний варіант; модуль визначення кількості приголосних фонем у вибірці; модуль визначення середніх частот груп приголосних фонем; модуль перевірки вибірок текстів на нормальність (критерій Пірсона); модуль визначення критеріїв Стюдента, Колмогорова-Смірнова та хі-квадрат. Побудована структура ґрунтується на модульному принципі, що забезпечує швидку модифікацію системи.

2. Розроблено алгоритм роботи та ПЗ системи диференціації фоностатистичних структур стилів англійської мови. Розроблене ПЗ є кросплатформним, оскільки розроблення системи здійснено на мові програмування Java. Для розроблення програми використано середовище програмування IntelliJIdeaCommunity 14. Дане середовище розробки компанії JetBrains містить широкий спектр інтегрованих інструментів, які цілком забезпечують розроблення програми. Додаткові технології та фреймворки не використовувались. Для написання програми використано такі стандартні бібліотеки Java як: swing – для написання користувацького інтерфейсу; awt – для додаткових функцій у користувацькому інтерфейсі, зокрема, опрацювання подій; I/O – для роботи з

файлами та текстами у них. Для спрощення написання математичних функцій використано спеціальну математичну бібліотеку Colt, яка спростила реалізацію формул дисперсії та критеріїв Стюдента, Колмогорова-Смірнова та хі-квадрат.

3. Описано особливості роботи та реалізації ПЗ. У результаті розроблення системи створено такі класи: `Main`, `TranscriptionProcessor`, `ConsonantProcessor`, `ConsonantUtils`, `StatisticProcessor`. У класі `Main` знаходиться пакет `flow`, який забезпечує завантаження файлів, послідовність виконання операцій та отримання результатів опрацювання. Для порівняння двох текстів використано HTTP протокол, через який здійснено завантаження двох файлів і вибір критерію перевірки гіпотези на однорідність вибірки. Клас `TranscriptionProcessor` відповідає за перетворення тексту у його транскрипційний варіант. Текст розбивається на масив слів. Для кожного слова перевіряється його наявність у базі даних. У програмі розроблено POST запити двох типів: `/process` і `/process/transcription`. Перший запит використовується при відсутності транскрипційного варіанту досліджуваного тексту, другий – при його наявності. Другий запит дає змогу скоротити час роботи програми. Якщо слово відсутнє, тоді за допомогою класу від `java.net.HttpURLConnection`, унаслідуваного від `java.net.URLConnection`, який підтримує з'єднання по мережевому протоколу HTTP, посилається запит на сайт транскрипційного перекладу <http://upodn.com/phon.php>. Для формування вибірки з приголосних фонем створено клас `ConsonantProcessor`. Вибірка обсягом у 51000 приголосних фонем формується вилученням з транскрипційних символів всіх нефонематичних символів і символів голосних фонем. Якщо вибірка не має достатнього обсягу, програма повідомляє про це. Вибірка ділиться на 51 порцію по 1000 фонем у кожній за допомогою функції `System.arraycopy()`, яка копіює створений масив символів швидше, ніж при використанні `for`-циклу. 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. Після проведення обчислень для кожної групи отримано таку структуру даних: `List<Map<ConsonantType, Long>>`. На основі отриманих результатів для кожної порції сформовано структуру даних `MultiValueMap<ConsonantType, Long>`. Це – структура

даних типу ключ-значення, у якій ключем є тип приголосного, а значенням – вибірка. Наступним кроком є перевірка двох вибірок на однорідність. Якщо вибірки неоднорідні, вони відрізняються авторським стилем. Програма перевіряє вибірки на однорідність за критеріями Стюдента, χ^2 і Колмогорова-Смірнова. Для вибору критерію, необхідно у полі *'statistical criterion'* вказати його назву. Клас `StatisticProcessor` відповідає за статистичне опрацювання вибірок та визначає критерій, який потрібно застосовувати. У цьому класі є метод `process`, який за параметри приймає назву критерію і вибірки, які зіставляються.

4. Проведено тестування розробленого ПЗ на прикладі розмежування англійських текстів публіцистичного стилю. Отримані результати дають змогу ствердити, що система працює правильно і коректно. За результатами тестування визначено значний ступінь дії фактора авторської манери викладу.

5. Наведено результати досліджень, зокрема: отримані числові дані з довірчою ймовірністю 0,95 уможливлують наступну характеристику досліджуваних стилів: при зіставленні розмовного стилю з поезією Байрона і Мура істотні розходження виявлено за методом гіпотез за 6–4 групами фонем; з художньою прозою Байрона за – 7–5 групами фонем; з драмою Шоу – за 4 групами фонем; з газетним стилем – за 6–3 групами фонем; з науковим стилем – за 7–4 групами фонем. При зіставленні розмовного стилю з драмою Шоу виявлено мінімальну (1 одиниця) різницю рангових показників, яку встановлено на основі зіставлення спадних рядів середніх частот групи сонорних фонем у випадку позиції фонем на початку слова, що підтверджує близькість між розмовним стилем і драмою Шоу. Незначну відстань драми Шоу від розмовного стилю (0,02) встановлено методом визначення відстаней між стилями для випадку позиції фонем на початку слова. Таким чином, встановлено середні істотні відмінності між розмовним стилем і драмою Шоу, що експериментально підтверджує наявність спільних елементів щоденно-побутової лексики у зіставлених вибірках.

При порівнянні функціональних стилів попарно між собою, встановлено певну близькість поезії Байрона і Мура (істотні розходження за 3–2 групами фонем за методом гіпотез та незначною різницею рангових показників – 2–3 одиниці за

методом ранжування), яка підтверджує схожість фонематичного наповнення у текстах приналежних до одного функціонального стилю та підстилю. В інших випадках попарного зіставлення стилів між собою встановлено істотні відмінності за більшістю груп приголосних фонем.

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 багатofакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу, який ґрунтується на визначенні загальної стилевої маркованості зіставлюваних текстів. Загальна стилева маркованість досліджуваного тексту передбачає встановлення істотних відмінностей у трьох зіставленнях (стилів, підстилів, текстів різних авторів) з довірчою ймовірністю 0,95, що уможливорює 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 уможливорює 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 ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Эйнштейн А. О науке / Собрание научных трудов. Москва : Наука, 1967. Т. 4. 600 с.
2. Liebich B. Panini. Lpz., 1891. 110 st.
3. Bourdon B. L'expression des emotions et des tendances dans le langage. Paris, 1892. 115 p.
4. Kaeding F. W. Das häufigkeitswörterbuch und die geläufigkeitsuntersuchungen // *Magazin für Stenographie XX*, 1899. S. 83–87.
5. Wang W. S-Y., Crawford J. C. Frequency studies of English consonants // *Language and Speech*, 1960. V. 3. Part 3. P. 131–139.
6. Humboldt W. Die sprachphilosophischen werke. Berlin, 1883. 232 s.
7. Соссюр Ф. де. Труды по языкознанию / пер. с франц. Москва : Прогресс, 1977. 695 с.
8. Программный труд выдающегося датского лингвиста Луи Ельмслева “Пролегомены к теории языка”. Москва, 2006. 248с.
9. Бодуэн де Куртенэ И. А. Избранные труды по общему языкознанию. Москва : Изд-во АН СССР, 1963. Т. 1. 384 с.
10. Щерба Л. В. Языковая система и речевая деятельность. Ленинград, 1974. 428 с.
11. Тимченко Є. Курс історії українського язика. Вступ і фонетика. Київ, 1927. URL: <https://uk.wikipedia.org/wiki/Фонема> (Last accessed: 03.01.2020).
12. Кочерган М. П. Загальне мовознавство. К: ВЦ Академія, 2008. 464с.
13. Солнцев В. М. Язык как системно-структурное образование. Москва : Наука, 1971. 292 с.
14. Мельничук А. С. Понятие системы и структуры языка. *Вопросы языкознания*. 1970. № 1. С. 19–32.
15. Coupe C., Marsico E., Pellegrino F. Structural complexity of phonological systems. *Phonology and Phonetics Series, Mouton de Gruyter*. Berlin, 2009. P. 141–170.
16. Hoole P., Bombien L., Pouplier M., Mooshammer C., Kuhnert B. Consonant clusters and structural complexity. Germany, Berlin : Walter de Gruyter, 2012. 409p.

17. Статистичні та структурні лінгвістичні моделі : Республіканський міжвідомчий збірник. Київ : Наукова думка, 1966. 161 с.
18. Ревзин И. И. Структура языка как моделирующей системы. Москва : Наука, 1978. 287 с.
19. Пиотровский Р. Г. Моделирование фонологических систем и методы их сравнения. Москва–Ленинград : Наука, 1966. 299 с.
20. Сегал Д. М. Основы фонологической статистики. Москва : Наука, 1972. 255 с.
21. Перебийніс В. С. Статистичні параметри стилів. Київ : Наукова думка, 1967. 240 с.
22. Кукушкина О.В., Поликарпов А.А., Хмелёв Д.В. Определение авторства текста с использованием буквенной и грамматической информации // *Проблемы передачи информации*. Москва, 2001. Т. 37. №2. С. 96–108.
23. Koppel M., Schler J., Argamon Sh. Authorship attribution: what's easy and what's hard? // *Computer Science*. 2013. 15p. DOI: 10.2139/ssrn.2274891.
24. Висоцька В. А., Шестакевич Т. В., Щербина Ю. М. Застосування породжувальних граматик для моделювання синтаксису речення // *Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі*. Львів, 2012. № 743. С. 175–190.
25. Висоцька В. А. Концептуальна модель процесу формування семантики речення природною мовою // *Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі*. Львів, 2014. № 805. С. 258–278.
26. Висоцька В. А. Особливості моделювання синтаксису речення слов'янських та германських мов за допомогою породжувальних контекстно-вільних граматик // *Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі*. Львів, 2015. № 814. С. 246–276.
27. Shestakevych T., Vysotska V., Chyrun L., Chyrun L. Modelling of semantics of natural language sentences using generative grammars // *Computer Science and Information Technologies (CSIT'2014): Proceedings of the IXth Scientific and Technical Conference*. Lviv, 18-22 November, 2014. P. 19–22.

28. Lupenko S., Orobchuk O., Mingtang Xu. Logical-structural models of verbal, formal, and machine-interpreted knowledge representation in integrative scientific medicine // *Advances in Intelligent Systems and Computing IV*. Springer, Natalia Shakhovska editor. 2019. P. 139–153.
29. Уильям Томсон, лорд Кельвин. URL: http://www.nsc.ru/win/elbib/data/show_page.dhtml?76+123 (Last accessed: 05.01.2020).
30. Буняковський В. Я. Основи математичної теорії ймовірностей. 1846. URL: https://uk.wikipedia.org/wiki/Буняковський_Віктор_Якович (Last accessed: 02.01.2020).
31. Zipf G. K. Relative frequency as a determinant of phonetic change // *Harvard Studies in Classical Philology*. Cambridge, 1929, N 40. P. 5–10.
32. Zipf G. K. Statistical methods and dynamic philology // *Language*. 1937. V. 13. P. 60–70.
33. Neophytou K., van Egmond M., Avrutin S. Zipf's law in aphasia across languages. A Comparison of English, Hungarian and Greek // *Journal of Quantitative Linguistics*. 13 Jan 2017. P. 178–196.
34. Herdan G. The advanced theory of language as choice and chance. Berlin, 1966. 142 p.
35. Бектаев К. Б., Пиотровский Р. Г. Математические методы в языкознании. Алма-Ата, 1973. Ч. 1. 281 с.
36. Бектаев К. Б., Пиотровский Р. Г. Математические методы в языкознании. Алма-Ата, 1974. Ч. 2. 260 с.
37. Köhler R., Altmann G. Aims and methods of quantitative linguistics // *Problemy kvantitativnoj lingvistiki*. Ćernivci: Ruta, 2005. P. 12–41.
38. Shyrovov V., Ostapova I., Nadutenko M., Verbynenko Yu. Ontologized lexicographical systems in modern terminography // *Cognitive Studies*, 2016. P. 90–99. DOI 10.11649/cs.2016.008.
39. Карпіловська Є. А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика : Донецьк: Юго-Восток, 2006. 187 с.

40. Дарчук Н. П. Комп'ютерна лінгвістика (автоматичне опрацювання тексту). Підручник. Київ : Видавничо-поліграфічний центр "Київський університет", 2008. 351 с.
41. Селіванова О. О. Сучасна лінгвістика. Напрями та проблеми. Полтава : Довкілля-К, 2008. 712 с.
42. Kornai A. *Mathematical linguistics*. Springer, 2008. 85 p.
43. Gries Th. S. *Statistics for linguistics with R. Mouton Textbook*, 2009. 335 p.
44. Висоцька В. А., Пасічник В. В., Щербина Ю. М., Шестакевич Т. В. Математична лінгвістика. Книга 1. Квантитативна лінгвістика : навч. посібник з грифом Міністерства освіти і науки, молоді та спорту України. Львів : видавництво «Новий світ-2000», 2012. 359 с.
45. Перебийніс В. С. Статистичні методи для лінгвістів. Вінниця : Нова Книга, 2013. 170 с.
46. Сегал Д. М. Статистическая однородность текста на фонологическом уровне в польском языке. *Структурная типология языков*. Москва, 1968. С. 85–93.
47. Deng Y. Some statistical properties of phonemes in standard Chinese // *Journal of Quantitative Linguistics*. Taylor & Francis Group. London, 2016. Vol. 23, issue 1. P. 30–48.
48. Koppel M., Schler J., Argamon Sh. Authorship attribution in the wild // *Language Resources and Evaluation*. 2011. Vol. 45. No. 1. P. 83–94. URL: <https://doi.org/10.1007/s10579-009-9111-2> (Last accessed: 13.02.2020).
49. Koppel M. Computational methods in authorship attribution // *Journal of the Association for Information Science and Technology*. USA, 2009. Vol. 60 (1). P. 9–26.
50. Koppel M., Schler J., Argamon Sh., Winter Ya. The “fundamental problem” of authorship attribution // *Electronic journal* 2012. Vol. 93. Issue 3. P. 284 – 291. DOI:10.1080/0013838X.2012.668794.
51. Boukhaled M. A., Ganascia J.-G. Using function words for authorship attribution: Bag-of-words vs. sequential rules // *Natural Language Processing and Cognitive Science : Proceedings*. 2015. P. 115–122.

52. Bisikalo O. V., Vysotska V. A. Sentence syntactic analysis application to keywords identification ukrainian texts // *Radio electronics computer science control. Zaporizhzhya*, 2016. № 3 (38). P. 54–65.
53. Lytvyn V. V., Bobyk I. O., Vysotska V. A. Application of algorithmic algebra system for grammatical analysis of symbolic computation expressions of propositional logic // *Radio Electronics Computer Science Control. Zaporizhzhya* : Zaporizhzhya National Technical University, 2016/4. Vol. 4 (39). P. 77–89.
54. Lytvyn V. Vysotska V., Pukach P., Bobyk I., Uhryn D. Development of a method for the recognition of author's style in the ukrainian language texts based on linguometry, stylemetry and glottochronology // *Eastern-European Journal of Enterprise Technologies*. ISSN 1729-3774. 2017. № 4/2 (88). P. 10–18.
55. Juola P. Future trends in authorship attribution // *Advances in Digital Forensics III*. Springer, 2007. P. 119–132.
56. Juola P. Authorship attribution, foundations and trends (r) in information retrieval // Boston–Delft, 2008. Vol. 1. No 3. P. 233–334.
57. Kapociute-Dzikiene J., Utka F., Sarkute L. Authorship attribution and author profiling of lithuanian literary texts // *Proceedings of the 5th Workshop on Balto-Slavic Natural Language Processing*. Hissac, Bulgaria, 2015. P. 96–105.
58. Stamatatos E. A survey of modern attribution methods // *Journal of the Association for Information Science and Technology*. 2009. Vol. 60. P. 538–556.
59. Stamatatos E., Daelemans W., Verhoeven B., Potthast M., Stein B., Juola P., Sanchez-Perez M. Overview of the author identification task at PAN 2014. P. 877 – 898 // *In Working Notes Papers of the CLEF 2014 Evaluation Labs, CEUR Workshop Proceedings*. CLEF. 2014.
60. Stamatatos E., Daelemans W., Verhoeven B., Juola P., Lopez Lopez A., Potthast M., Stein B. : Overview of the author identification task at PAN 2015 // *In Working Notes Papers of the CLEF 2015 Evaluation Labs, CEUR Workshop Proceedings*. CLEF and CEUR-WS.org. ISSN 1613-0073. 2015. 17p.

61. Stamatatos E. Authorship attribution using text distortion // *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Vol. 1, 2017. P. 1138–1149.
62. Yule G. U. Kendal M. G. Wstep do teorii statystyki. Warszawa : PWN, 1966. 684 p.
63. Miles J. Style and Proportion. Boston, 1967. 158 p.
64. Guiraud P. Problemes et methodes de la statistique linguistique. Paris, 1960. 148 p.
65. Guiraud P. Zagadnienia i metody statystyki jezykoznawczej. Warszawa : PWN, 1966. 159 s.
66. Кожина М. Н. К основаниям функциональной стилистики. Пермь, 1968. 251 с.
67. Мацько Л. І. Стиль як основна категорія лінгво-стилістики // *Семантика і структура східнослов'янських мов: До 200-річчя з дня народження І. Даля*. Київ, 2001. С. 262–268.
68. Гриднева Л. М. Розподіл голосних, приголосних, пропусків у сучасному українському мовленні // *Статистичні та структурні лінгвістичні моделі*. Київ, 1966. С. 45–53.
69. Головин Б. Н. Вопросы статистической стилистики. Київ : Наукова думка, 1974. 330 с.
70. Тулдава Ю. А. О теоретико-методологических основах квантитативно-системного анализа лексики // *Лингвистика текста и стилистика*. Тарту, 1981. С. 114–133.
71. Guillén Nieto V., Vargas Sierra C., Pardiño Juan M., Martínez Barco P., Suárez Cueto A. Exploring state-of-the-art software for forensic authorship identification / *International Journal of English Studies*, Vol. 8. No 1. 2008. P. 1–28. URL: <https://revistas.um.es/ijes/article/view/49071>. (Last Accessed: 23.10. 2019).
72. Baayen R. H. Word frequency distributions, Springer, Netherlands, eBook. 193p. ISBN 978-94-010-0844-0. DOI 10.1007/978-94-010-0844-0. 2001
73. Ganapathi Raju N.V., Sadhvi Ch., Tejaswini P., Mounica Y. Style based authorship attribution on english editorial documents // *International Journal of Computer Applications*. India, 2017. Vol. 159. No 4. P. 5 – 8.

74. Petmanson T. Authorship identification using correlations of frequent features. Notebook for PAN at CLEF. Institute of Computer Science, University of Tartu. 2013. 20p. URL: <http://ceur-ws.org/Vol-1179/CLEF2013wn-PAN-Petmanson2013.pdf>. (Last Accessed: 20.10. 2019).
75. Волошиновська І. А. Сильова, тематична й авторська атрибуція наукових і художніх текстів (на матеріалі англійської, німецької та української мов): автореферат дис. канд. філол. наук: 10.02.15. Наук. керівник Т. С. Толчєєва. М-во науки і освіти України. Нац. пед. ун-т ім М. П. Драгоманова. Київ, 2018. 18с.
76. Granichin O., Kizhaeva N., Shalymov D., Volkovich Z. Writing style determination using the KNN text model // *Proceedings of the 2015 IEEE International Symposium on Intelligent Control*. Sydney, Australia. Sept. 21-23, 2015. P. 900–905.
77. Wilson A. Units and constituency in prosodic analysis // A Quantitative Assessment. *Journal of Quantitative Linguistics*. 24 Nov. 2016. P. 163–177.
78. Authorship attribution software. URL: <https://neoneuro.com/ru/products/authorship-attribution> (Last accessed: 13.01.2020).
79. Unplag. URL: <https://ua.unplag.com/features-for-corporate-clients/> (Last accessed: 22.02.2020).
80. What is the best tool or software for stylometry. URL: <https://www.quora.com/What-is-the-best-tool-or-software-for-stylometry> (Last accessed: 23.08.2020).
81. Online authorship attribution tool. URL: <http://www.aicbt.com/authorship-attribution/online-software/> (Last accessed: 1.03.2020).
82. Evaluating language variations in EVL Labs. URL: <https://evllabs.com/> (Last accessed: 25.08.2020).
83. The java graphical authorship attribution program. URL: <https://github.com/evllabs/JGAAP#java-graphical-authorship-attribution-program> (Last accessed: 24.08.2020).
84. Гнеденко Б. В. Курс теорії ймовірностей. Київ : Київський університет, 2010. 464 с.
85. Сеньо П. С. Теорія ймовірностей та математичної статистики : підручник. Київ : Центр навчальної літератури, 2004. 448 с.

86. Пушак Я. С., Лозовий Б. Л. Теорія імовірностей і елементи математичної статистики. Львів : Українська академія друкарства, 2006. 264 с.
87. Рабик В. М. Основи теорії імовірностей: навч. посібник. Львів : Магнолія плюс, 2006. 176 с.
88. Forbes C., Evans M., Hastings N. Peacock B. Statistical distributions. New Jersey: A John Wiley & Sons, 2011. 231p.
89. Jones M. C. Student's simplest distribution // *Journal of the Royal Statistical Society. Series D*, 51. 2002. P. 41-49.
90. Oztuna D., Elhan AH., Tuccar E. Investigation of four different normality tests in terms of type 1 error rate and power under different distributions // *Turkish Journal of Medical Sciences*. 2006. 36(3):171–6.
91. Patel J. K., Read C. B. Handbook of the normal distribution. New York: Marcel Dekker, 1982. 344p.
92. Thode H. J. Testing for normality. New York: Marcel Dekker. 2002. 368p.
93. Митропольский А. К. Техника статистических вычислений. Москва : Наука, 1971. 576 с.
94. Пиотровский Р. Г., Бектаев К. Б., Пиотровская А. А. Математическая лингвистика. Москва : Высшая школа, 1977. 383 с.
95. Kolmogorov A. N. Foundations of the theory of probability. Chelsea Publishing, 1950. 95p.
96. Mathematics. Its content, methods and meaning. Edited by A. D. Aleksandrov, A. N. Kolmogorov and M. A. Lavrent'ev. Dover Publications, Inc., Mineola, NY, 1999. 372p.
97. Ивченко Г. И., Медведев Ю.И. Математическая статистика. Москва: Высш. шк., 1984. 248с.
98. Steinskog D. J. A cautionary note on the use of the Kolmogorov-Smirnov test for normality. *American Meteor Soc.* 2007. 135:1151–7.
99. Rozanov Y. A. Probability theory: A concise course (Dover Books on Mathematics). Dover Publications, 1977. 162p.

100. Турчин В. М. Теорія ймовірностей і математична статистика. Дніпропетровськ: ІМА-прес, 2014. 294 с.
101. Sinai Ya. G. Probability theory. An introductory course. Springer. 1992. 138p.
102. Snedecor G. W., Cochran W. G. Statistical methods. Iowa; Iowa State Press. USA. 1989. 233p.
103. Probability theory and mathematical statistics. Proceedings of the Fifth Japan-USSR Symposium. Editors: Watanabe Sh., Prokhorov Yu. Japan, Kyoto. July 8-14. 1986. 590p.
104. Мартыненко Г. Я., Чебанов С. В. Текст как объект и предмет стилеметрии // *Квантитативные аспекты системной организации текста*. Материалы межвузовского семинара. Тбилиси, 1987. С. 83–86.
105. Мартыненко Г. Я. Основы стилеметрии. Ленинград : Изд-во ЛГУ, 1988. 173 с.
106. Мартыненко Г. Я., Чебанов С. В. Классификационные задачи стилеметрии // *Квантитативная лингвистика и автоматический анализ текста*. Тарту, 1988. С. 119–136.
107. Трубецкой Н. С. Избранные труды по филологии. Москва : Прогресс, 1987. 559 с.
108. Васильев В. А., Катанская А. Р., Лукина Н. Д., Маслова Л. П., Торсуева Е. И. Фонетика английского языка. Нормативный курс. Москва : Высш. школа, 1980. 256 с.
109. Зиндер Л. Р., Строева Т. В. К вопросу о применении статистики в языкознании // *Вопросы языкознания*. 1968. № 6. С. 120–123.
110. Batyuk A., Voityshyn V., Verhun V. “Software architecture design of the real-time processes monitoring platform” // *Proceeding of the Second International Conference on Data Stream Mining & Processing (DSMP)*. Lviv, 2018. P 98–101.
111. Niemeyer P., Knudsen J. “Learning Java.” *O'Reilly&Associates*, 2000. Inc. 716.
112. 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.

113. Free french travel phrasebook. URL: <https://www.mosalingua.com/en/free-travel-language-phrasebooks/free-french-travel-phrasebook/> (Last accessed: 24.02.2020).
114. Burrows J. Delta: a measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*. 2002. 17/3. P. 267–287.
115. Хомицька І. Ю. Диференціація функціональних стилів англійської мови за ранговими показниками середніх частот груп фонем // *Гуманітарний вісник. Серія: Іноземна філологія*. Черкаси, 2003. Число сьоме. С. 193–196.
116. Тамбовцев Ю. О., Хомицька І. Ю. Статистичні параметри стилів англійської мови // *Іноземна філологія*. Львів, 1993. Вип. 106. С. 54–59.
117. Тамбовцев Ю. О., Хомицька І. Ю. Розрізнення функціональних стилів на фонематичному рівні // *Іноземна філологія*. Львів, 1995. Вип. 108. С. 84–88.
118. Хомицька І. Ю. До проблеми взаємодії розмовного та художнього функціональних стилів англійської мови // *Гуманітарний вісник. Серія: Іноземна філологія*. Черкаси, 2008. Число дванадцяте. Т. 2. С. 144–148.
119. Хомицька І. Ю. Застосування методу гіпотез для визначення взаємодії літературної різновидності розмовного стилю та підстилю драми художнього стилю англійської мови // *Збірник наукових праць. Людина. Комп'ютер. Комунікація*. Львів, 2008. С. 180–181.
120. Хомицька І. Ю. Залежність стилерозмежувальної потужності груп фонем англійської мови від позиції фонем в слові // *Іноземна філологія*. Львів, 2001. Вип. 112. С. 192–196.
121. 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.
122. 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.

123. 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.
124. Хомицька І. Ю. Деякі фоностатистичні особливості функціональних стилів англійської мови // *Гуманітарний вісник. Серія: Іноземна філологія*. Черкаси, 2004. Число восьме. С. 321–324.
125. Хомицька І. Ю. Моделювання статистичних структур функціональних стилів англійської мови // *Наукові записки. Серія: філологічні науки (мовознавство)*. Кіровоград, 2015. Вип. 136. С. 209–213.
126. 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.
127. 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.
128. Selections from Byron. Moscow : Progress Publishers, 1973. 37 p.
129. The prisoner of Chillon and selections from other works of Lord Byron. Bielefeld und Leipzig, 1930. 320 p.
130. Coleridge S. T. Verse and prose. – Moscow : Progress Publishers, 1981. 447 p..
131. Moore Th. Selected verse. Moscow, 1981. 544 p.
132. Moore Th. Selected verse. Moscow, 1986. 542 p.
133. Хомицька І. Ю. До проблеми впливу творчої індивідуальності автора на диференціацію функціональних стилів англійської мови на фонологічному рівні // *Гуманітарний вісник. Серія: Іноземна філологія*. Черкаси, 2001. Число п'яте. С. 238–241.
134. Хомицька І. Ю. Концепт «Сум» у творах Дж. Г. Байрона «Манфред», «Шильйонський в'язень», «Корсар» // *Гуманітарний вісник. Серія: Іноземна філологія*. Черкаси, 2010. Число чотирнадцяте. Т. 1. С. 138–143.

135. Хомицька І. Ю. Застосування методів математичної статистики для диференціації текстів поеми Дж. Г. Байрона «Гяур» та поезії Т. Мура «Ірландські мелодії» // *Збірник наукових праць. Людина. Комп'ютер. Комунікація*. Львів, 2010. С. 102–103.
136. Хомицька І. Ю. Вербалізація концепта боротьба у поемі Дж. Г. Байрона «Корсар» // *Наукові записки. Серія: Філологічні науки (мовознавство)*. Кіровоград, 2012. Вип. 104 (2). С. 291–294.
137. Хомицька І. Ю. Фоностатистичні особливості мовної картини світу в поемі Дж. Г. Байрона «Корсар» // *Наукові записки. Серія: Філологічні науки (мовознавство)*. Кіровоград, 2013. Вип. 119. С. 177–180.
138. Хомицька І. Ю. Специфіка фонематичного вираження мовної картини світу в поемі Дж. Г. Байрона «Шильйонський в'язень» // *Збірник наукових праць Людина. Комп'ютер. Комунікація*. Львів, 2013. С. 181–184.
139. Хомицька І. Ю., Теслюк В. М. Статистичний аналіз англійських поетичних текстів // *Науковий вісник. НЛТУ України*. Львів, 2015. Вип. 25.2. С. 350–356.
140. Хомицька І. Ю. Концепт сум в поемі Дж. Г. Байрона «Мазепа» // *Наукові записки. Серія: Філологічні науки (мовознавство)*. Кіровоград, 2014. Вип. 127. С. 336–339.
141. Khomytska I. Phonostatistical peculiarities of linguistic world image in poem lara by g.g. Byron // *French Journal of Scientific and Educational Research*. Paris: Paris University Press. 2014. P. 894 – 903.
142. Khomytska I. Phonostatistical structure of poem the bride of abydos by g.g. Byron in phonostatistical system of english styles // *American Journal of Scientific and Educational Research*. New York: Columbia Press. 2014. P. 400-408.
143. 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.
144. Byron L. Childe Harold's pilgrimage. London, 1892. 320 p.

145. 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.
146. 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.
147. Short M. Exploring the language of poems, plays and prose. London: Longman. 1996. 399p.
148. Khomytska I., Teslyuk V., Holovatyu 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.
149. Selections from Shaw. A fearless champion of the truth. Moscow : Progress Publishers, 1977. 420 p.
150. Birch D. The language of drama. Series: Language of Literature. Palgrave, 1991. 184p.
151. Хомицька І. Ю. Залежність середніх частот вживання груп приголосних фонем від типів мовленнєвих жанрів у драмі Б. Шоу «Другий острів Джона Буля» // *Наукові записки. Серія: Філологічні науки (мовознавство): У 2 ч.* Кіровоград, 2011. Вип. 96 (2). С. 326–329.
152. Хомицька І. Ю. Залежність середніх частот вживання груп приголосних фонем від типів мовленнєвих актів у драмі Б. Шоу «Другий острів Джона Буля» // *Англістика та американістика*. Дніпропетровськ, 2011. Вип. 8. С. 118–126.
153. 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.
154. Carter R., Nash W. Seeing through language. Great Britain : Basil Blackwell, 1990. 688 p.

155. "The Daily Telegraph". URL: <https://www.telegraph.co.uk/> (Last accessed: 25.02.2020).
156. Reah D. *The language of newspapers (Intertext)*. Routledge, 2002. 142p.
157. Хомицька І. Ю. До проблеми взаємодії газетного та художнього функціональних стилів англійської мови // *Гуманітарний вісник. Серія: Іноземна філологія*. Черкаси, 2007. Число одинадцяте. Т. 2. С. 513–519.
158. Хомицька І. Ю. До проблеми взаємодії газетного та розмовного функціональних стилів англійської мови // *Гуманітарний вісник. Серія: Іноземна філологія*. Черкаси, 2006. Число десяте. Т. 2. С. 409–413.
159. Хомицька І. Ю. До проблеми взаємодії газетного та наукового функціональних стилів англійської мови // *Гуманітарний вісник. Серія: Іноземна філологія*. Черкаси, 2005. Число дев'яте. С. 324–327.
160. 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.
161. Kulchytskyi I., Shandruk U. The quantitative research of scientific texts at the symbolic level // *Computational linguistics and intelligent systems*. Lviv : Lviv Polytechnic National University, 25 – 27 June. 2018. Vol. 2. P. 71–80.
162. Crystal D., Davy D. *Investigating english style*. Great Britain. Longman : London, 1969. 215 p.
163. Левченко О. П. *Науковий стиль : культура мовлення*. Львів : Видавництво Львівської політехніки. 2012. 204с.
164. *Acta physica Polonica* 2018. URL: <http://www.journals4free.com/link.jsp?l=6610578> (Last accessed: 9.03.2020).
165. 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.

166. Bhargava M., Mehndiratta P., Asawa K. Stylometric analysis for authorship attribution on twitter // *Proceedings of the Second International Conference on Big Data Analytics*. Vol. 8302. 2013. P. 37–47. https://doi.org/10.1007/978-3-319-03689-2_3
167. Zhao Y., Zobel J. Searching with style: Authorship attribution in classic literature // *Proceedings of the thirtieth Australasian conference on Computer science*. Australian Computer Society. 2007. Inc, vol. 62. P. 59–68.
168. 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.
169. 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.
170. 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.
171. 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.
172. Wuthering heights by Emily Bronte online. URL: <https://www.wuthering-heights.co.uk/wh/novel/html/index> (Last accessed: 12.03.2020).
173. Vanity fair by William Makepeace Thackeray - Read Online. URL: <http://www.literaturepage.com/read/vanity-fair.html> (Last accessed: 15.03.2020).
174. Casas R. M. Systems for the phonetic transcription of English: Theory and Texts. Series: Linguistic Insights. Peter Lang AG, 2011. 280p.

175. The classification of english consonants. URL: <https://studfile.net/preview/5544077/>
(Last accessed: 26.08.2020).
176. The Freedom Papers | Tavaana. URL: <https://tavaana.org/en/library-asset-collection-new/press-freedom> (Last accessed: 29.02.2020).
177. Freedom Paper No. 1: Free & Independent Media – Air University. URL: <http://igmlnet.uohyd.ac.in:8000/InfoUSA/politics/freedom/freedom1.htm> (Last accessed: 10.03.2020).
178. Category: speeches by Donald Trump – Wikipedia. URL: https://en.wikipedia.org/wiki/Category:Speeches_by_Donald_Trump (Last accessed: 11.03.2020).
179. Barack Obama 450 + Speeches – Transcripts, Audio, Video. URL: <https://www.americanrhetoric.com/barackobamaspeeches.htm> (Last accessed: 13.03.2020).
180. 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, 2, P. 23–26.
181. Khomytska I., Teslyuk V., Kryvinska N., Bazylevych I. Software-based approach towards automated authorship acknowledgement—chi-square test on one consonant group // Electronics. July 2020. Vol. 7:1138. DOI:10.3390/electronics9071138.
182. 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.
183. 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. AISC 871 (Selected Papers from the International Conference on Computer Science and Information Technologies, CSIT 2018, September 11-14, Lviv, Ukraine), Springer, January 2019. P.105–118. DOI: 10.1007/978-3-030-01069-0_8.

184. 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.

Додатки

Додатки А.1 – А.64. Результати порівняння текстів за методом гіпотез

Таблиця А.1

Порівняння розмовного стилю з поезією Дж. Г. Байрона

Групи фонем	Байрон		Розмовний стиль		S	t	$2Q$	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	134,3	2774,26	131,9	7611,48	13,16	0,72	> 20%	Випадковий
Передньо-язикові	421,8	6401,44	362,9	32500,3	25,46	9,11	< 0,1%	Істотний
Середньо-язикові	6,9	305,58	18,6	5175,36	9,56	4,82	< 0,1%	Істотний
Задньо-язикові	64,3	2940,01	72,6	5157,36	11,62	2,81	< 1,0%	Істотний
Носові	83,6	3649,29	76,8	4202,84	11,44	2,34	< 5%	Істотний
Сонорні	224,7	6546,76	226,9	14575,5	18,76	0,46	> 50%	Випадковий
Щілинні	190,3	11210,59	158,9	7948,71	17,87	6,92	< 0,1%	Істотний
Проривні	202,9	5599,58	226,5	5725,74	13,74	6,76	< 0,1%	Істотний

Таблиця А.2

Порівняння розмовного стилю з поезією Дж. Г. Байрона
фонем на початку слова

Групи фонем	Байрон		Розмовний стиль		S	t	2Q	Тип значення величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	76,0	3147,00	72,00	5077,57	11,71	1,34	> 10%	Випадковий
Передньо-язикові	111,3	2452,39	100,0	7930,97	13,16	3,38	< 2%	Істотний
Середньо-язикові	2,7	147,99	11,6	3409,36	7,70	4,55	< 0,1%	Істотний
Задньо-язикові	34,8	2664,84	31,3	2880,77	9,61	2,25	< 5%	Істотний
Носові	9,8	731,44	5,7	336,77	4,22	3,83	< 0,1%	Істотний
Сонорні	63,5	5183,75	68,4	5383,10	13,27	1,45	> 5%	Випадковий
Щілинні	97,6	4201,36	69,3	3813,94	11,56	9,64	< 0,1%	Істотний
Проривні	66,8	2232,84	78,2	5504,19	11,36	3,95	< 0,1%	Істотний

Таблиця А.3

Порівняння розмовного стилю з поезією Дж. Г. Байрона

Фонемі в кінці слова

Групи фонем	Байрон		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	22,1	907,51	23,5	1943,74	5,70	0,97	> 20%	Випадковий
Передньо-язикові	165,1	6359,71	128,7	4594,77	13,51	10,60	< 0,1%	Істотний
Середньо-язикові	–	–	–	–	–	–	–	–
Задньо-язикові	12,4	501,36	18,4	993,36	4,99	4,73	< 0,1%	Істотний
Носові	33,0	1511,00	34,4	1913,68	7,55	0,73	> 20%	Випадковий
Сонорні	55,3	1973,99	53,6	1557,36	7,67	0,87	> 20%	Випадковий
Щілинні	66,1	2099,51	49,8	3069,42	9,28	6,91	< 0,1%	Істотний
Проривні	79,0	6844,00	67,6	2373,68	12,39	3,62	< 0,1%	Істотний

Таблиця А.4

Порівняння розмовного стилю з поезією Т. Мура

Групи фонем	Мур		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	137,9	4156,56	131,9	7611,48	14,00	1,69	> 5%	Випадковий
Передньо-язикові	425,0	8178,00	362,9	32500,3	26,04	9,39	< 0.1%	Істотний
Середньо-язикові	5,9	143,58	18,6	5175,36	9,42	5,31	< 0.1%	Істотний
Задньо-язикові	59,3	3242,26	72,6	5157,36	11,83	4,43	< 0.1%	Істотний
Носові	82,9	1902,71	76,8	4202,84	10,09	2,38	< 5%	Істотний
Сонорні	233,9	4890,01	226,9	14575,5	18,01	1,53	> 10%	Випадковий
Щілинні	210,3	8529,18	158,9	7948,71	16,57	12,21	< 0.1%	Істотний
Проривні	182,7	10670	226,5	5725,74	1653	10,43	< 0.1%	Істотний

Таблиця А.5

Порівняння розмовного стилю з поезією Т. Мура

Фонем на початку слова

Групи фонем	Мур		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	71,8	1842,24	72,0	5077,57	10,74	0,07	> 80%	Випадковий
Передньо-язикові	115,7	3421,99	100,0	7930,97	13,76	4,49	< 0,1%	Істотний
Середньо-язикові	2,3	130,39	11,6	3409,36	7,68	4,77	< 0.1%	Істотний
Задньо-язикові	31,8	1128,24	31,3	2880,77	8,17	0,24	> 80%	Випадковий
Носові	7,7	355,99	5,7	336,77	3,40	2,32	< 5%	Істотний
Сонорні	63,1	2009,91	68,4	5383,10	11,10	1,88	> 5%	Випадковий
Щілинні	101,9	6095,91	69,3	3813,94	12,85	9,99	< 0,1%	Істотний
Проривні	57,1	2217,51	78,2	5504,19	11,34	7,32	< 0,1%	Істотний

Таблиця А.6

Порівняння розмовного стилю з поезією Т. Мура

Фонемі в кінці слова

Групи фонем	Мур		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	26,4	1519,10	23,5	1943,74	7,60	1,50	> 10%	Випадковий
Передньо-язикові	161,1	3853,91	128,7	4594,77	11,87	10,75	< 0,1%	Істотний
Середньо-язикові	–	–	–	–	–	–	–	–
Задньо-язикові	14,8	925,44	18,4	993,36	5,66	2,51	< 2%	Істотний
Носові	34,4	931,56	34,4	1913,68	6,89	0,00	100%	Випадковий
Сонорні	57,0	1495,00	53,6	1557,36	7,13	1,88	> 5%	Випадковий
Щілинні	71,0	3223,00	49,8	3069,12	10,24	8,15	< 0,1%	Істотний
Проривні	74,5	3255,75	67,6	2373,68	9,69	2,80	< 1%	Істотний

Таблиця А.7

Порівняння розмовного стилю з художньою прозою Дж. Г. Байрона

Групи фонем	Байрон		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	139,7	3253,22	131,9	7611,48	13,46	2,28	< 5%	Істотний
Передньо-язикові	402,7	7542,39	362,9	32500,3	25,83	6,07	< 0,1%	Істотний
Середньо-язикові	6,7	313,99	18,6	5175,36	9,56	4,90	< 0,1%	Істотний
Задньо-язикові	60,9	2243,51	72,6	5157,36	11,11	4,15	< 0,1%	Істотний
Носові	80,0	3005,00	76,8	4202,84	10,96	1,15	> 20%	Випадковий
Сонорні	210,4	5463,56	226,9	14575,5	18,28	3,55	< 0,1%	Істотний
Щілинні	204,2	10770,24	158,9	7948,71	17,66	10,10	< 0,1%	Істотний
Проривні	194,1	12960,3	226,5	5725,74	17,65	7,23	< 0,1%	Істотний

Таблиця А.8

Порівняння розмовного стилю з художньою прозою Дж. Г. Байрона

Фонемі на початку слова

Групи фонем	Байрон		Розмовний стиль		S	t	$2Q$	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	64,0	1058,00	72,0	5077,57	10,11	3,11	< 0,5	Істотний
Передньо-язикові	93,2	13225,44	100,0	7930,97	18,78	1,43	> 10%	Випадковий
Середньо-язикові	1,3	49,99	11,6	3409,36	7,59	5,34	< 0,1%	Істотний
Задньо-язикові	30,6	2169,56	31,3	2880,77	9,17	0,30	> 50%	Випадковий
Носові	6,7	382,39	5,7	336,77	3,46	1,14	> 20%	Випадковий
Сонорні	48,4	2865,56	68,4	5383,10	11,72	6,72	< 0,1%	Істотний
Щілинні	83,9	10265,51	69,3	3813,94	15,32	3,75	< 0,1%	Істотний
Проривні	56,8	1875,44	78,2	5504,19	11,09	7,60	< 0,1%	Істотний

Таблиця А.9

Порівняння розмовного стилю з художньою прозою Дж. Г. Байрона

Фонемі в кінці слова

Групи фонем	Байрон		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	29,2	1353,42	23,5	1943,74	7,41	3,03	< 0,5%	Істотний
Передньо-язикові	135,2	4720,84	128,7	4594,77	12,46	2,05	< 5%	Істотний
Середньо-язикові	–	–	–	–	–	–	–	–
Задньо-язикові	8,6	407,59	18,4	993,36	4,83	7,98	< 0,1%	Істотний
Носові	30,7	1022,79	34,4	1913,68	7,00	2,08	< 5%	Істотний
Сонорні	48,0	1879,00	53,6	1557,36	7,57	2,91	0,5%	Істотний
Щілинні	65,0	6553,56	49,8	3069,42	12,66	4,73	< 0,1%	Істотний
Проривні	58,1	2194,71	67,6	2373,68	8,73	4,29	< 0,1%	Істотний

Таблиця А.10

Порівняння розмовного стилю з драмою Б. Шоу

Групи фонем	Шоу		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	121,1	2992,71	131,9	7611,48	13,29	3,20	< 0,5%	Істотний
Передньо-язикові	390,2	9504,84	362,9	32500,3	26,46	4,06	< 0,1%	Істотний
Середньо-язикові	18,6	1557,56	18,6	5175,36	10,59	0,00	100%	Випадковий
Задньо-язикові	68,5	2707,75	72,6	5157,36	11,45	1,41	> 10%	Випадковий
Носові	92,2	6098,84	76,8	4202,84	13,10	4,63	< 0,1%	Істотний
Сонорні	234,2	5611,44	226,9	14575,5	18,34	1,57	> 10 %	Випадковий
Щілинні	162,6	5351,56	158,9	7948,71	14,89	0,98	> 20%	Випадковий
Проривні	202,6	5055,36	226,5	5725,74	13,40	7,02	< 0,1%	Істотний

Таблиця А.11

**Порівняння розмовного стилю з драмою Б. Шоу
фонем на початку слова**

Групи фонем	Шоу		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	59,1	3807,51	72,0	5077,57	12,17	4,17	< 0,1%	Істотний
Передньо-язикові	95,8	1735,44	100,0	7930,97	12,69	1,30	20%	Випадковий
Середньо-язикові	16,4	1743,56	11,6	3409,36	9,27	2,04	< 5%	Істотний
Задньо-язикові	33,6	1619,16	31,3	2880,77	8,66	1,05	> 20%	Випадковий
Носові	11,5	441,75	5,7	336,77	3,60	6,34	< 0,1%	Істотний
Сонорні	71,4	3009,16	68,4	5383,10	11,83	1,00	> 20%	Випадковий
Щілинні	71,6	3747,56	69,3	3813,94	11,23	0,81	> 20%	Випадковий
Проривні	62,2	1316,24	78,2	5504,19	10,66	5,91	< 0,1%	Істотний

Таблиця А.12

**Порівняння розмовного стилю з драмою Б. Шоу
фонем в кінці слова**

Групи фонем	Шоу		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	26,2	742,84	23,5	1943,74	6,69	1,59	> 10%	Випадковий
Передньо-язикові	142,5	1247,75	128,7	4594,77	9,87	5,51	< 0,1%	Істотний
Середньо-язикові	–	–	–	–	–	–	–	–
Задньо-язикові	15,4	633,36	18,4	993,36	5,21	2,27	< 5%	Істотний
Носові	35,2	811,44	34,4	1913,68	6,74	0,47	> 50%	Випадковий
Сонорні	60,7	1740,39	53,6	1557,36	7,41	3,77	< 0,1%	Істотний
Щілинні	48,3	1919,99	49,8	3069,42	9,12	0,65	> 50%	Випадковий
Проривні	74,4	2069,16	67,6	2373,68	8,60	3,11	< 0,5%	Істотний

Таблиця А.13

Порівняння розмовного і газетного стилів

Групи фонем	Газетний стиль		Розмовний стиль		S	t	$2Q$	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	130,1	6565,51	131,9	7611,48	15,37	0,46	> 50%	Випадковий
Передньо-язикові	406,1	4991,76	362,9	32500,3	25,00	6,93	< 0,1%	Істотний
Середньо-язикові	7,3	1929,59	18,6	5175,36	10,88	4,09	< 0,1%	Істотний
Задньо-язикові	64,7	2702,39	72,6	5157,36	11,45	2,72	< 1%	Істотний
Носові	82,3	1450,79	76,8	4202,84	9,71	2,23	< 5%	Істотний
Сонорні	225,9	4517,51	226,9	14575,5	17,84	0,22	> 80%	Випадковий
Щілинні	183,4	8051,36	158,9	7948,71	12,78	7,55	< 0,1%	Істотний
Проривні	204,5	4025,75	226,5	5725,74	12,75	6,79	< 0,1%	Істотний

Таблиця А.14

**Порівняння розмовного і газетного стилів
фонем на початку слова**

Групи фонем	Газетний стиль		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	60,4	5795,16	72,0	577,57	13,46	3,39	< 0,2%	Істотний
Передньо-язикові	90,3	5333,09	100,0	7930,97	14,87	2,57	< 2%	Істотний
Середньо-язикові	4,5	733	11,6	3409,36	8,31	3,36	< 0,2%	Істотний
Задньо-язикові	25,1	1089,91	31,3	2880,77	8,13	3,00	< 0,5%	Істотний
Носові	5,3	384,39	5,7	336,77	3,47	0,45	> 50%	Випадковий
Сонорні	50,2	1738,84	68,4	5383,10	10,89	6,58	< 0,1%	Істотний
Щілинні	69,5	2357,75	69,3	3813,94	10,14	0,08	> 80%	Випадковий
Проривні	60,5	5259,75	78,2	5504,19	13,39	5,20	< 0,1%	Істотний

Таблиця А.15

**Порівняння розмовного і газетного стилів
фонем в кінці слова**

Групи фонем	Газетний стиль		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	20,5	563,75	23,5	1943,74	6,46	1,83	> 5%	Випадковий
Передньо-язикові	134,9	5073,11	128,7	4594,77	12,69	1,92	> 5%	Випадковий
Середньо-язикові	–	–	–	–	–	–	–	–
Задньо-язикові	14,6	473,56	18,4	993,36	4,94	3,03	< 0,5%	Істотний
Носові	33	682	34,4	1913,68	6,58	0,84	> 20%	Випадковий
Сонорні	54,3	1752,39	53,6	1557,36	7,43	0,37	> 50%	Випадковий
Щілинні	58,4	2055,36	49,8	3069,42	9,24	3,66	< 0,1%	Істотний
Проривні	56,0	3470	67,6	2373,68	9,87	4,63	< 0,1%	Істотний

Таблиця А.16

Порівняння розмовного і наукового стилів

Групи фонем	Науковий стиль		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	123,5	5465,75	131,9	7611,48	14,76	2,24	< 5%	Істотний
Передньо-язикові	425,1	11442,71	362,9	32500,3	27,06	9,05	< 0,1%	Істотний
Середньо-язикові	8,4	419,36	18,6	5175,36	9,66	4,16	< 0,1%	Істотний
Задньо-язикові	60,9	2282,71	72,6	5157,36	11,14	4,14	< 0,1%	Істотний
Носові	87,6	3125,91	76,8	4202,84	11,05	3,95	< 0,1%	Істотний
Сонорні	220,4	6461,16	226,9	14575,5	18,72	1,37	> 10%	Випадковий
Щілинні	186,8	2872,24	158,9	7948,71	13,43	8,18	< 0,1%	Істотний
Проривні	211,1	4202,71	226,5	5725,74	12,86	4,71	< 0,1%	Істотний

Таблиця А.17

Порівняння розмовного і наукового стилів
Фонем на початку слова

Групи фонем	Науковий стиль		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	52,0	2711,00	72,0	5077,57	11,39	6,91	< 0,1%	Істотний
Передньо-язикові	80,4	3005,36	100,0	7930,97	13,50	5,72	< 0,1%	Істотний
Середньо-язикові	1,5	57,75	11,6	3409,36	7,60	5,44	< 0,1%	Істотний
Задньо-язикові	21,0	1297,00	31,3	2880,77	8,34	4,86	< 0,1%	Істотний
Носові	2,9	170,71	5,7	336,77	2,91	3,79	< 0,1%	Істотний
Сонорні	30,4	687,16	68,4	5383,10	10,06	14,87	< 0,1%	Істотний
Щілинні	70,0	2138,00	69,3	3813,94	9,96	0,28	> 50%	Випадковий
Проривні	54,9	3164,71	78,2	5504,19	12,02	7,63	< 0,1%	Істотний

Таблиця А.18

Порівняння розмовного і наукового стилів
Фонем в кінці слова

Групи фонем	Науковий стиль		Розмовний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	18,5	683,75	23,5	1943,74	6,62	2,98	< 0,5%	Істотний
Передньо-язикові	125,8	2447,44	128,7	4594,77	10,83	1,05	> 20%	Випадковий
Середньо-язикові	–	–	–	–	–	–	–	–
Задньо-язикові	10,1	432,71	18,4	993,36	4,88	6,70	< 0,1%	Істотний
Носові	37,0	1889,00	34,4	1913,68	7,96	1,29	> 20%	Випадковий
Сонорні	54,2	3338,24	53,6	1557,36	9,03	0,26	> 50%	Випадковий
Щілинні	56,5	14,23,75	49,8	3069,42	8,65	3,05	< 0,5%	Істотний
Проривні	43,3	1220,79	67,6	2373,68	7,74	2,36	< 0,1%	Істотний

Таблиця А.19

Порівняння поезії і художньої прози Дж. Г. Байрона

Групи фонем	Байрон (поезія)		Байрон (худ. проза)		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	134,3	2774,26	139,7	3253,22	10,02	2,08	< 5%	Істотний
Передньо-язикові	421,8	6401,44	402,7	7542,39	15,24	4,93	< 0,1%	Істотний
Середньо-язикові	6,9	305,58	6,7	313,99	3,21	0,25	> 20%	Випадковий
Задньо-язикові	64,3	2940,01	60,9	2243,51	9,29	1,44	> 10%	Випадковий
Носові	83,6	3649,29	80,0	3005,00	10,53	1,35	> 10%	Випадковий
Сонорні	224,7	6546,76	210,4	5463,56	14,15	3,98	< 0,1%	Істотний
Щілинні	190,3	11210,59	204,2	10770,24	19,14	2,96	< 1%	Істотний
Проривні	202,9	5599,58	194,1	12960,31	17,59	1,97	> 5%	Випадковий

Таблиця А.20

Порівняння поезії і художньої прози Дж. Г. Байрона
Фонем на початку слова

Групи фонем	Байрон (поезія)		Байрон (худ. проза)		s	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	76,0	3147,00	64,0	1058,00	8,37	5,64	< 0,1%	Істотний
Передньо-язикові	111,3	2452,39	93,2	13225,44	16,16	9,41	< 0,1%	Істотний
Середньо-язикові	2,7	147,99	1,3	49,99	1,82	3,03	< 0,5%	Істотний
Задньо-язикові	36,8	2664,84	30,6	2169,56	8,98	2,72	< 1%	Істотний
Носові	9,8	731,44	6,7	382,39	4,31	2,83	< 1%	Істотний
Сонорні	63,5	5183,75	48,4	2865,56	11,58	5,13	< 0,1%	Істотний
Щілинні	97,6	3201,36	83,9	10265,51	15,53	3,47	< 0,1%	Істотний
Проривні	66,8	2232,84	56,8	1875,44	8,27	4,76	< 0,1%	Істотний

Таблиця А.21

Порівняння поезії і художньої прози Байрона
Фонем в кінці слова

Групи фонем	Байрон (поезія)		Байрон (худ. проза)		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	22,1	907,51	29,2	1353,42	6,14	4,55	< 0,1%	Істотний
Передньо-язикові	165,1	6358,71	135,2	4720,88	13,59	8,66	< 0,1%	Істотний
Середньо-язикові	–	–	–	–	–	–	–	–
Задньо-язикові	12,4	501,36	8,6	407,56	3,89	3,89	< 0,1%	Істотний
Носові	33,0	1511,00	30,7	1022,79	6,50	1,39	> 10%	Випадковий
Сонорні	55,3	1973,99	48,0	1879,00	8,01	3,59	< 0,1%	Істотний
Щілинні	66,1	2099,51	65,0	6553,56	12,01	0,36	> 50%	Випадковий
Проривні	79,0	6844,00	58,1	2194,71	12,27	6,70	< 0,1%	Істотний

Таблиця А.22

Порівняння драми Б. Шоу з поезією Дж. Г. Байрона

Групи фонем	Шоу		Байрон		S	t	$2Q$	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	121,1	2992,71	134,3	2774,26	9,80	5,30	< 0,1%	Істотний
Передньо-язикові	390,2	9504,84	421,8	6401,44	16,28	7,64	< 0,1%	Істотний
Середньо-язикові	18,6	1557,56	6,9	305,58	5,57	8,27	< 0,1%	Істотний
Задньо-язикові	68,5	2707,75	64,3	2940,01	9,70	1,70	> 5%	Випадковий
Носові	92,2	6098,84	83,6	3649,29	12,75	2,66	1%	Істотний
Сонорні	162,6	5351,56	190,3	11210,59	16,61	6,57	< 0.1%	Істотний
Щілинні	202,6	5055,36	202,9	5599,58	13,33	0,09	> 80%	Випадковий
Проривні	234,2	5611,44	224,7	6546,76	14,24	2,63	< 1%	Істотний

Таблиця А.23

Порівняння драми Б. Шоу з поезією Дж. Г. Байрона
Фонем на початку слова

Групи фонем	Шоу		Байрон		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	59,1	3807,51	76,0	3147,00	10,77	6,18	< 0,1%	Істотний
Передньо-язикові	95,8	1735,44	111,3	2452,39	8,35	7,03	< 0,1%	Істотний
Середньо-язикові	16,4	1743,56	2,7	147,99	5,61	9,61	< 0,1%	Істотний
Задньо-язикові	33,6	1619,16	36,8	2664,84	8,45	1,49	> 10%	Випадковий
Носові	11,5	441,75	9,8	731,44	4,42	1,51	> 10%	Випадковий
Сонорні	71,6	3747,56	97,6	4201,36	11,51	8,89	< 0,1%	Істотний
Щілинні	62,2	1316,24	66,8	2232,84	7,69	2,35	< 5%	Істотний
Проривні	71,4	3009,16	63,5	5183,75	11,69	2,66	< 1%	Істотний

Таблиця А.24

Порівняння драми Б. Шоу з поезією Дж. Г. Байрона
Фонем в кінці слова

Групи фонем	Шоу		Байрон		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	26,2	742,84	22,1	907,51	5,24	3,08	< 0,5%	Істотний
Передньо-язикові	142,5	1247,75	165,1	6358,71	11,26	7,90	< 0,1%	Істотний
Середньо-язикові	–	–	–	–	–	–	–	–
Задньо-язикові	15,4	633,36	12,4	501,36	4,35	2,72	< 1%	Істотний
Носові	35,2	811,44	33,0	1511,0	6,22	1,39	> 10%	Випадковий
Сонорні	60,7	1740,39	55,3	1973,99	7,86	2,70	< 1%	Істотний
Щілинні	49,3	1919,99	66,1	2099,51	8,18	8,08	< 0,1%	Істотний
Проривні	74,4	2069,16	79,0	6844,00	12,19	1,49	> 10%	Випадковий

Таблиця А.25

Порівняння поезії Дж. Г. Байрона з газетним стилем

Групи фонем	Байрон		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	134,3	2774,26	130,1	6565,51	12,48	1,32	> 10%	Випадковий
Передньо-язикові	421,8	6401,44	406,4	4991,76	13,78	4,40	< 0,1%	Істотний
Середньо-язикові	6,9	305,58	7,3	1929,59	6,10	0,26	> 50%	Випадковий
Задньо-язикові	64,3	2940,01	64,7	2702,39	9,70	0,16	> 80%	Випадковий
Носові	83,6	3649,29	82,3	1450,79	9,22	0,56	> 50%	Випадковий
Сонорні	224,7	6546,76	225,9	4517,51	13,58	0,34	> 50%	Випадковий
Щілинні	190,3	11210,59	183,4	1851,36	14,75	1,84	> 5%	Випадковий
Проривні	202,9	5599,58	204,5	4025,75	12,66	0,50	> 50%	Випадковий

Таблиця А.26

Порівняння поезії Дж. Г. Байрона з газетним стилем
Фонем на початку слова

Групи фонем	Байрон		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	76,0	3147,00	60,4	5795,16	12,21	5,03	< 0,1%	Істотний
Передньо-язикові	111,3	2452,39	90,3	5333,09	11,39	7,26	< 0,1%	Істотний
Середньо-язикові	2,7	147,99	4,5	733	3,83	1,85	> 5%	Випадковий
Задньо-язикові	36,8	2664,84	25,1	1089,91	7,91	5,82	< 0,1%	Істотний
Носові	9,8	731,44	5,3	384,39	4,31	4,11	< 0,1%	Істотний
Сонорні	63,5	5183,75	50,2	1738,84	10,74	4,88	< 0,1%	Істотний
Щілинні	97,6	4201,36	69,5	2357,75	10,46	10,58	< 0,1%	Істотний
Проривні	66,8	2232,84	60,5	5259,75	11,17	2,22	< 5%	Істотний

Таблиця А.27

Порівняння поезії Дж. Г. Байрона з газетним стилем
Фонемі в кінці слова

Групи фонем	Байрон		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	22,1	901,51	20,5	563,75	4,95	1,27	> 20%	Випадковий
Передньо-язикові	165,1	6359,71	134,9	5073,11	1370	8,62	< 0,1%	Істотний
Середньо-язикові	–	–	–	–	–	–	–	–
Задньо-язикові	12,4	501,36	14,6	473,56	4,03	2,14	< 5%	Істотний
Носові	33,0	1511,00	33	682	6,04	–	100%	Випадковий
Сонорні	55,3	1973,99	54,3	1752,39	7,88	0,50	> 50%	Випадковий
Щілинні	66,1	2099,51	58,4	2055,36	8,32	3,64	< 0,1%	Істотний
Проривні	79,0	6844,00	56,0	3470	13,11	6,91	< 0,1%	Істотний

Таблиця А.28

Порівняння поезії Дж. Г. Байрона з науковим стилем

Групи фонем	Байрон		Науковий стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	134,3	2774,26	123,5	5465,75	11,72	3,63	< 0,1%	Істотний
Передньо-язикові	421,8	6401,44	425,1	11442,71	17,25	0,75	> 20%	Випадковий
Середньо-язикові	6,9	305,58	8,4	419,36	3,48	1,70	> 5%	Випадковий
Задньо-язикові	64,3	2940,01	60,9	2282,71	9,33	1,43	> 10%	Випадковий
Носові	83,6	3649,29	87,9	3125,91	10,63	1,59	> 10%	Випадковий
Сонорні	224,7	6546,76	220,4	6461,16	14,72	1,15	> 20%	Випадковий
Щілинні	190,3	11210,59	186,8	2872,24	15,32	0,90	> 20%	Випадковий
Проривні	202,9	5599,69	211,1	4202,71	12,78	2,53	< 2%	Істотний

Таблиця А.29

Порівняння поезії Дж. Г. Байрона з науковим стилем
Фонем на початку слова

Групи фонем	Байрон		Науковий стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	76,0	3147,00	52,0	2711,00	9,88	9,56	< 0,1%	Істотний
Передньо-язикові	111,3	2452,39	80,4	3005,36	9,54	12,76	< 0,1%	Істотний
Середньо-язикові	2,7	147,99	1,5	57,75	1,85	2,55	< 2%	Істотний
Задньо-язикові	36,8	2664,84	21,0	1297,00	8,13	7,66	< 0,1%	Істотний
Носові	9,8	731,44	2,9	170,71	3,88	7,01	< 0,1%	Істотний
Сонорні	63,5	5183,75	30,4	687,16	9,89	13,17	< 0,1%	Істотний
Щілинні	97,6	4201,36	70,0	2138,00	10,28	10,57	< 0,1%	Істотний
Проривні	66,8	2232,84	54,9	3164,71	9,48	4,94	< 0,1%	Істотний

Таблиця А.30

Порівняння поезії Дж. Г. Байрона з науковим стилем
Фонеми в кінці слова

Групи фонем	Байрон		Науковий стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	22,1	907,51	18,5	583,75	4,99	2,84	< 1%	Істотний
Передньо-язикові	165,1	6358,71	125,8	2447,44	12,11	12,77	< 0,1%	Істотний
Середньо-язикові	–	–	–	–	–	–	–	–
Задньо-язикові	12,4	501,36	10,1	432,71	3,95	2,29	< 5%	Істотний
Носові	33,0	1511,00	37,0	1889,00	7,53	2,09	< 5%	Істотний
Сонорні	55,3	1973,99	54,2	3338,24	9,41	0,46	> 50%	Випадковий
Щілинні	66,1	2099,51	56,5	1423,75	7,66	4,93	< 0,1%	Істотний
Проривні	79,0	6844,00	43,3	1220,79	11,59	12,12	< 0,1%	Істотний

Таблиця А.31

Порівняння художньої прози Байрона з поезією Т. Мура

Групи фонем	Байрон		Мур		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	139,7	3253,22	137,9	4156,56	11,11	0,64	> 20%	Випадковий
Передньо-язикові	402,7	7542,39	425,0	8178,00	16,19	5,42	< 0,1%	Істотний
Середньо-язикові	6,7	313,99	5,9	143,58	2,76	1,14	> 20%	Випадковий
Задньо-язикові	60,9	2243,51	59,3	3242,26	9,56	0,66	> 20%	Випадковий
Носові	80,0	3005,00	82,9	1902,71	9,04	1,26	> 20%	Випадковий
Сонорні	210,4	5463,56	233,9	4890,01	13,14	7,04	< 0,1%	Істотний
Щілинні	204,2	10770,24	210,3	8529,19	17,93	1,34	> 10%	Випадковий
Зімкнені	194,1	12960,31	182,7	10670,01	19,85	2,26	< 5%	Істотний

Таблиця А.32

*Порівняння художньої прози Дж.Г. Байрона з поезією Т. Мура
Фонема на початку слова*

Групи фонем	Байрон		Мур		S	t	2Q	Тип значення величини $\bar{X}_1 - \bar{X}_2$
	\bar{X}	$\sum (X_i - \bar{X})^2$	\bar{X}	$\sum (X_i - \bar{X})^2$				
Губні	64,0	1058,00	71,8	1842,24	6,95	4,42	<0,1%	Істотний
Передньо-язикові	93,2	13225,44	115,7	3421,99	16,66	5,31	<0,1%	Істотний
Середньо-язикові	1,3	39,99	2,3	130,39	1,73	2,27	<5%	Істотний
Задньо-язикові	30,6	2169,56	31,8	1128,24	7,41	0,64	>50%	Випадковий
Носові	6,7	382,39	7,7	355,99	3,51	1,12	>20%	Випадковий
Сонорні	48,4	2865,56	63,1	2009,91	9,01	6,42	<0,1%	Істотний
Щілинні	83,9	10265,51	101,9	6095,91	16,51	4,29	<0,1%	Істотний
Зімкнені	56,8	1875,44	57,1	2217,51	8,26	0,14	>80%	Випадковий

Таблиця А.33

*Порівняння художньої прози Дж.Г. Байрона з поезією Т. Мура
Фонема в кінці слова*

Групи фонем	Байрон		Мур		S	t	2Q	Тип значення величини $\bar{X}_1 - \bar{X}_2$
	\bar{X}	$\sum (X_i - \bar{X})^2$	\bar{X}	$\sum (X_i - \bar{X})^2$				
Губні	29,2	1353,42	26,4	1519,16	6,92	1,59	>10%	Випадковий
Передньо-язикові	135,2	4720,84	161,1	3853,91	11,95	8,53	<0,1%	Істотний
Середньо-язикові	-	-	-	-	-	-	-	-
Задньо-язикові	8,6	407,56	14,8	925,44	4,71	5,18	<0,1%	Істотний
Носові	30,7	1022,79	34,4	931,56	5,71	2,55	<2%	Істотний
Сонорні	48,0	1879,00	57,0	1495,00	7,50	4,73	<0,1%	Істотний
Щілинні	65,0	6553,56	71,0	3223,00	12,76	1,85	>5%	Випадковий
Зімкнені	58,1	2194,71	74,5	3255,75	9,53	6,77	<0,1%	Істотний

Таблиця А.34

Порівняння драми Б. Шоу з поезією Т. Мура

Групи фонем	Шоу		Мур		S	t	2Q	Тип значення величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	121,1	2992,71	137,9	4156,56	10,92	6,06	< 0,1%	Істотний
Передньо-язикові	390,2	9504,84	425,0	8178,00	17,17	7,98	< 0,1%	Істотний
Середньо-язикові	18,6	1557,56	5,9	143,58	5,32	9,40	< 0,1%	Істотний
Задньо-язикові	68,5	2707,75	59,3	3242,26	9,96	3,64	< 0,1%	Істотний
Носові	92,2	6098,84	82,9	1902,71	11,55	3,17	< 0,5%	Істотний
Щілинні	162,6	5351,56	210,3	8529,18	15,21	12,35	< 0,1%	Істотний
Зімкнені	202,6	5055,36	182,7	10670,00	16,19	4,84	< 0,1%	Істотний
Сонорні	234,2	5611,44	233,9	4890,01	13,23	0,09	> 80%	Випадковий

Таблиця А.35

Порівняння драми Б. Шоу з поезією Т. Мура
Фонем на початку слова

Групи фонем	Шоу		Мур		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	59,1	3807,51	71,8	1842,24	9,70	5,15	< 0,1%	Істотний
Передньо-язикові	95,8	1735,44	115,7	3421,99	9,27	8,45	< 0,1%	Істотний
Середньо-язикові	16,4	1743,56	2,3	130,39	5,59	9,93	< 0,1%	Істотний
Задньо-язикові	33,6	1619,16	31,8	1128,24	6,77	1,05	> 20%	Випадковий
Носові	11,5	441,75	7,7	355,99	3,65	4.10	< 0,1%	Істотний
Сонорні	71,4	3009,16	63,1	2009,91	9,15	3.57	< 0,1%	Істотний
Щілинні	71,6	3747,56	101,9	6095,91	12,81	9.31	< 0,1%	Істотний
Зімкнені	62,2	1316,24	57,1	2217,51	7,67	2,62	< 2%	Істотний

Таблиця А.36

Порівняння драми Б. Шоу з поезією Т. Мура
Фонем в кінці слова

Групи фонем	Шоу		Мур		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	26,2	742,84	26,4	1519,16	6,14	0,13	> 80%	Випадковий
Передньо-язикові	142,5	1247,75	161,1	3853,91	9,22	7,94	< 0,1%	Істотний
Середньо-язикові	–	–	–	–				
Задньо-язикові	15,4	633,36	14,8	925,44	5,10	0,46	> 50%	Випадковий
Носові	35,2	811,44	34,4	931,56	5,39	0,58	> 50%	Випадковий
Сонорні	60,7	1740,39	57,0	1495,00	7,34	1,98	> 5%	Випадковий
Щілинні	49,3	1919,99	71,0	3223,00	9,26	9,23	< 0,1%	Істотний
Зімкнені	74,4	2069,16	74,5	3255,75	9,42	0,04	> 80%	Випадковий

Таблиця А.37

Порівняння поезії Т. Мура з газетним стилем

Групи фонем	Мур		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	137,9	4156,56	130,1	6565,51	13,36	2,30	< 5%	Істотний
Передньо-язикові	425,0	8178,00	406,4	4991,76	14,82	4,94	< 0,1%	Істотний
Середньо-язикові	5,9	143,58	7,3	1929,59	5,88	0,93	> 20%	Випадковий
Задньо-язикові	59,3	3242,26	64,7	2702,39	9,95	2,14	< 5%	Істотний
Носові	82,9	1902,71	82,3	1450,79	7,48	0,32	> 50%	Випадковий
Сонорні	233,9	4899,01	225,9	4517,51	12,52	2,52	< 2%	Істотний
Щілинні	210,3	8529,19	183,4	1851,36	13,15	8,05	< 0,1%	Істотний
Зімкнені	182,7	10670,01	204,5	4025,75	15,65	5,48	< 0,1%	Істотний

Таблиця А.38

Порівняння поезії Т. Мура з газетним стилем

Фонем на початку слова

Групи фонем	Мур		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	71,8	1842,24	60,4	5795,16	11,28	3,98	< 0,1%	Істотний
Передньо-язикові	115,7	3421,99	90,3	5333,09	12,08	8,28	< 0,1%	Істотний
Середньо-язикові	2,3	130,39	4,5	7,33	3,79	2,28	< 5%	Істотний
Задньо-язикові	31,8	1128,24	25,1	1089,91	6,08	4,34	< 0,1%	Істотний
Носові	7,7	355,99	5,3	384,39	3,51	2,69	< 1%	Істотний
Сонорні	63,1	2009,91	50,2	1738,84	7,90	6,42	< 0,1%	Істотний
Щілинні	101,9	6095,91	69,5	2357,75	11,86	10,76	< 0,1%	Істотний
Зімкнені	57,1	2217,51	60,5	5259,75	11,16	1,2	> 20%	Випадковий

Таблиця А.39

Порівняння поезії Т. Мура з газетним стилем

Фонемі в кінці слова слова

Групи фонем	Мур		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	26,4	1519,16	20,5	563,75	5,89	3,94	< 0,1%	Істотний
Передньо-язикові	161,1	3853,91	134,9	5073,11	12,20	8,45	< 0,1%	Істотний
Середньо-язикові	–	–	–	–				
Задньо-язикові	14,8	925,44	14,6	473,56	4,82	0,16	> 80%	Випадковий
Носові	34,4	931,56	33	682	5,18	1,06	> 20%	Випадковий
Сонорні	57,0	1495,00	54,3	1752,39	7,36	1,44	> 10%	Випадковий
Щілинні	71,0	3223,00	58,4	2055,36	9,38	5,28	< 0,1%	Істотний
Зімкнені	74,5	3255,75	56,0	3470	10,58	6,88	< 0,1%	Істотний

Таблиця А.40

Порівняння поезії Т. Мура з науковим стилем

Групи фонем	Мур		Науковий стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	137,9	4156,56	123,5	5465,75	12,66	4,48	< 0,1%	Істотний
Передньо-язикові	425,0	8178,00	425,1	11442,71	18,08	0,02	> 20%	Випадковий
Середньо-язикові	5,9	143,58	8,4	419,36	3,06	3,22	< 0,5%	Істотний
Задньо-язикові	59,3	3242,26	60,9	2282,71	9,6	0,66	> 20%	Випадковий
Носові	82,9	1902,71	87,9	3125,91	9,15	2,15	< 5%	Істотний
Сонорні	233,9	4890,01	220,4	6461,16	13,75	3,87	< 0,1%	Істотний
Щілинні	210,3	8529,19	186,8	2872,24	13,78	6,71	< 0,1%	Істотний
Зімкнені	182,7	10670,01	211,1	4202,71	15,74	7,10	< 0,1%	Істотний

Таблиця А.41

Порівняння поезії Т. Мура з науковим стилем

Фонемі на початку слова

Групи фонем	Мур		Науковий стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	71,8	1842,24	52,0	2711,0	8,71	8,95	< 0,1%	Істотний
Передньо-язикові	115,7	3421,99	80,4	3005,36	10,35	13,43	< 0,1%	Істотний
Середньо-язикові	2,3	130,39	1,5	57,75	1,77	1,78	> 5%	Випадковий
Задньо-язикові	31,8	1128,24	21,0	1297,00	6,36	6,69	< 0,1%	Істотний
Носові	7,7	355,99	2,9	170,71	2,96	6,38	< 0,1%	Істотний
Сонорні	63,1	2009,91	30,4	687,16	6,70	19,20	< 0,1%	Істотний
Щілинні	101,9	6095,91	70,0	2138,00	11,71	10,72	< 0,1%	Істотний
Зімкнені	57,1	2217,51	54,9	3164,71	9,47	0,91	> 20%	Випадковий

Таблиця А.42

Порівняння поезії Т. Мура з науковим стилем

Фонемі в кінці слова

Групи фонем	Мур		Науковий стиль		S	t	$2Q$	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	26,4	1519,16	18,5	583,75	5,92	5,25	< 0,1%	Істотний
Передньо-язикові	161,1	3853,91	125,8	2447,44	10,25	13,56	< 0,1%	Істотний
Середньо-язикові	–	–	–	–				Істотний
Задньо-язикові	14,8	925,44	10,1	432,71	4,76	3,06	< 0,5%	Випадковий
Носові	34,4	931,56	37,0	1889,00	6,86	1,49	> 10%	Випадковий
Сонорні	57,0	1495,00	54,2	3338,24	8,98	1,23	> 20%	Істотний
Щілинні	71,0	3223,00	56,5	1423,75	8,80	6,49	< 0,1%	Істотний
Зімкнені	74,5	3255,75	43,3	1220,79	8,64	14,22	< 0,1%	Істотний

Таблиця А.43

Порівняння поезії Дж.Г. Байрона з поезією Т. Мура

Групи фонем	Байрон		Мур		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	134,3	2774,26	137,9	4156,56	10,75	1,318	≈20%	Випадковий
Передньо-язикові	421,8	6401,44	425,0	8176,00	15,59	0,808	>20%	Випадковий
Середньо-язикові	6,9	305,58	5,9	143,58	2,74	1,437	>10%	Випадковий
Задньо-язикові	64,3	2940,01	59,3	3242,26	10,15	1,939	>5%	Випадковий
Носові	83,6	3649,29	82,9	1902,71	9,62	0,286	>20%	Випадковий
Сонорні	224,7	6546,76	233,9	4890,01	13,81	2,623	≈1%	Істотний
Щілинні	190,3	11210,59	210,3	8529,19	18,14	4,341	<0,1%	Істотний
Зімкнені	202,9	5599,58	182,7	10670,01	16,47	4,829	<0,1%	Істотний

Таблиця А.44

Порівняння поезії Дж.Г. Байрона з поезією Т. Мура

Фонемі на початку слова

Групи фонем	Байрон		Мур		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	76,0	3147,00	71,8	1842,34	9,12	1,81	>5%	Випадковий
Передньо-язикові	111,3	2452,39	115,7	3421,99	9,89	1,75	>5%	Випадковий
Середньо-язикові	2,7	147,99	2,3	130,39	2,15	0,73	≈50%	Випадковий
Задньо-язикові	36,8	2664,84	31,8	1128,24	7,95	2,48	<2%	Істотний
Носові	9,8	731,44	7,7	355,99	4,26	1,94	>5%	Випадковий
Сонорні	63,5	5183,75	63,1	2009,91	10,95	0,14	>80%	Випадковий
Щілинні	97,6	4201,36	101,9	6095,91	13,10	1,29	≈20%	Випадковий
Зімкнені	66,8	2232,84	57,1	2217,51	8,61	4,43	<0,1%	Істотний

Таблиця А.45

Порівняння поезії Дж.Г. Байрона з поезією Т. Мура

Фонемі в кінці слова

Групи фонем	Байрон		Мур		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	22,1	907,51	26,4	1519,16	6,36	2,66	1%	Істотний
Передньо-язикові	165,1	6358,71	161,1	3853,91	13,05	1,21	>20%	Випадковий
Середньо-язикові	-	-	-	-	-	-	-	Випадковий
Задньо-язикові	12,4	501,36	14,8	926,44	4,88	1,94	>5%	Випадковий
Носові	33,0	1511,00	34,4	931,56	6,38	0,86	>20%	Випадковий
Сонорні	55,3	1973,99	57,0	1495,00	7,60	0,88	>20%	Випадковий
Щілинні	66,1	2099,61	71,0	3223,00	9,42	2,05	<5%	Істотний
Зімкнені	79,0	6844,0	74,5	3255,75	12,97	1,37	>10%	Випадковий

Таблиця А.46

Порівняння драми Б. Шоу з художньою прозою Дж. Г. Байрона

Групи фонем	Шоу		Байрон		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	121,1	2992,71	139,7	3253,22	10,20	7,18	< 0,1%	Істотний
Передньо-язикові	390,2	9504,84	402,7	7542,39	16,86	2,92	< 0,5%	Істотний
Середньо-язикові	18,6	1557,56	6,7	313,99	5,59	8,38	< 0,1%	Істотний
Задньо-язикові	68,5	2707,75	60,9	2243,51	9,08	3,30	< 0,2%	Істотний
Носові	92,2	6098,84	80,0	3005,00	12,32	3,90	< 0,1%	Істотний
Щілинні	162,6	5351,56	204,2	10770,24	16,39	10,00	< 0,1%	Істотний
Зімкнені	202,6	5055,36	194,1	12960,30	17,33	1,93	> 5%	Випадковий
Сонорні	234,2	5611,44	210,4	5463,56	13,59	6,89	< 0,1%	Істотний

Таблиця А.47

Порівняння драми Б. Шоу з художньою прозою Дж. Г. Байрона

Фонемі на початку слова

Групи фонем	Шоу		Байрон		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	59,1	3807,51	64,0	1058,00	9,01	2,14	< 5%	Істотний
Передньо-язикові	95,8	1735,44	93,2	13225,44	15,79	0,65	> 50%	Випадковий
Середньо-язикові	16,4	1743,56	1,3	49,99	5,47	10,87	< 0,1%	Істотний
Задньо-язикові	33,6	1619,16	30,6	2169,56	7,95	1,48	> 10%	Випадковий
Носові	11,5	441,75	6,7	382,39	3,71	5,10	< 0,1%	Істотний
Сонорні	71,4	3009,16	48,4	2865,56	9,90	9,15	< 0,1%	Істотний
Щілинні	71,6	3747,56	83,9	10265,51	15,28	3,17	< 0,5%	Істотний
Зімкнені	62,2	1316,24	56,8	1875,44	7,29	2,91	0,5%	Істотний

Таблиця А.48

Порівняння драми Б. Шоу з художньою прозою Дж. Г. Байрона

Фонемі в кінці слова

Групи фонем	Шоу		Байрон		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	26,2	742,84	29,2	1353,42	5,91	2,00	5%	Випадковий
Передньо-язикові	142,5	1247,75	135,2	4720,84	9,98	2,76	< 0,1%	Істотний
Середньо-язикові	–	–	–	–				Істотний
Задньо-язикові	15,4	633,36	8,6	407,56	4,17	6,43	< 0,1%	Істотний
Носові	35,2	811,44	30,7	1022,79	5,53	3,20	< 0,5%	Істотний
Сонорні	60,7	1740,39	48,0	1879,00	7,77	6,44	< 0,1%	Істотний
Щілинні	49,3	1919,99	65,0	6553,56	11,88	5,20	< 0,1%	Істотний
Зімкнені	74,4	20,69,16	58,1	2194,71	8,43	7,61	< 0,1%	Істотний

Таблиця А.49

Порівняння художньої прози Дж. Г. Байрона з газетним стилем

Групи фонем	Байрон		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	139,7	3253,22	130,1	6565,51	12,80	2,95	< 0,5%	Істотний
Передньо-язикові	402,7	7542,39	406,4	4991,76	14,45	1,01	> 20%	Випадковий
Середньо-язикові	6,7	313,99	7,3	1929,59	6,11	0,38	> 50%	Випадковий
Задньо-язикові	60,9	2243,51	64,7	2702,39	9,08	1,64	> 5%	Випадковий
Носові	80,0	3005,00	82,3	1450,79	8,62	1,05	> 20%	Випадковий
Сонорні	210,4	5463,56	225,9	4517,51	12,90	4,73	< 0,1%	Істотний
Щілинні	204,2	10770,24	183,4	1851,36	14,50	5,64	< 0,1%	Істотний
Зімкнені	194,1	12960,31	204,5	4025,75	16,82	2,43	< 2%	Істотний

Таблиця А.50

Порівняння художньої прози Дж. Г. Байрона з газетним стилем

Фонеми на початку слова

Групи фонем	Байрон		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	64,0	1058,00	60,4	5795,16	10,68	1,32	> 10%	Випадковий
Передньо-язикові	93,2	13295,44	90,3	5333,09	17,58	0,64	> 50%	Випадковий
Середньо-язикові	1,3	49,99	4,5	733,00	3,61	3,48	< 0,1%	Істотний
Задньо-язикові	30,6	2169,56	25,1	1089,91	7,37	2,94	< 0,5%	Істотний
Носові	6,7	382,39	5,3	384,39	3,57	1,54	> 10%	Випадковий
Сонорні	48,4	2865,56	50,2	1738,84	67,86	0,10	> 80%	Випадковий
Щілинні	83,9	10265,51	69,5	2357,75	14,50	3,91	< 0,1%	Істотний
Зімкнені	56,8	1875,44	60,5	5259,75	10,91	1,34	> 10%	Випадковий

Таблиця А.51

Порівняння художньої прози Дж. Г. Байрона з газетним стилем

Фонемі в кінці слова

Групи фонем	Байрон		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	29,2	1353,42	20,5	563,75	5,65	6,06	< 0,1%	Істотний
Передньо-язикові	135,2	4720,84	134,9	5073,11	12,78	0,09	> 80%	Випадковий
Середньо-язикові	–	–	–	–				
Задньо-язикові	8,6	407,56	14,6	473,56	3,83	6,17	< 0,1%	Істотний
Носові	30,7	1022,79	33	682	5,33	1,70	> 5%	Випадковий
Сонорні	48,0	1879,00	54,3	1752,39	7,78	3,18	< 0,5%	Істотний
Щілинні	65,0	6553,56	58,4	2055,36	11,98	2,16	< 5%	Істотний
Зімкнені	58,1	2194,71	56,0	3470	9,72	0,85	> 20%	Випадковий

Таблиця А.52

Порівняння художньої прози Дж. Г. Байрона з науковим стилем

Групи фонем	Байрон		Науковий стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	139,7	3253,22	123,5	5465,75	12,5	5,29	< 0,1%	Істотний
Передньо-язикові	402,7	7542,39	425,1	11442,71	17,79	4,96	< 0,1%	Істотний
Середньо-язикові	6,7	313,99	8,4	419,36	3,50	1,91	5%	Випадковий
Задньо-язикові	60,9	2243,51	60,9	2282,71	8,69	0,00	100%	Випадковий
Носові	80,0	3005,00	87,9	3125,91	10,11	3,08	< 0,5%	Істотний
Сонорні	210,4	5463,56	220,4	6461,16	14,10	2,79	< 1%	Істотний
Щілинні	204,2	10770,24	186,8	2872,24	15,05	4,54	< 0,1%	Істотний
Зімкнені	194,1	12960,31	211,1	4202,71	16,91	3,96	< 0,1%	Істотний

Таблиця А.53

Порівняння художньої прози Дж. Г. Байрона з науковим стилем

Фонемі на початку слова

Групи фонем	Байрон		Науковий стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	64,0	1058,00	52,0	2711,00	7,93	5,96	< 0,1%	Істотний
Передньо-язикові	93,2	13225,44	80,4	3005,36	16,45	3,06	< 0,5%	Істотний
Середньо-язикові	1,3	49,99	1,5	57,75	1,34	0,59	50%	Випадковий
Задньо-язикові	30,6	2169,56	21,0	1297,00	7,60	4,97	< 0,1%	Істотний
Носові	6,7	382,39	2,9	170,71	3,04	4,93	< 0,1%	Істотний
Сонорні	48,4	2865,56	30,4	687,16	7,69	9,21	< 0,1%	Істотний
Щілинні	33,9	10265,51	70,0	2138,00	14,38	3,81	< 0,1%	Істотний
Зімкнені	56,8	1875,44	54,9	3164,71	9,17	0,82	> 20%	Випадковий

Таблиця А.54

Порівняння художньої прози Дж. Г. Байрона з науковим стилем

Фонемі в кінці слова

Групи фонем	Байрон		Науковий стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	29,2	1353,42	18,5	583,75	5,68	7,41	< 0,1%	Істотний
Передньо-язикові	135,2	4720,84	125,8	2447,44	10,93	3,39	< 0,2%	Істотний
Середньо-язикові	–	–	–	–				
Задньо-язикові	8,6	407,56	10,1	432,71	3,74	1,58	> 10%	Випадковий
Носові	30,7	1022,79	37,0	1889,00	6,97	3,56	< 0,1%	Істотний
Сонорні	48,0	1879,00	54,2	3338,24	9,32	2,62	< 2%	Істотний
Щілинні	65,0	6553,56	56,5	1423,75	11,53	2,90	< 1%	Істотний
Зімкнені	58,1	2194,71	43,3	1220,79	7,54	7,72	< 0,1%	Істотний

Таблиця А.55

Порівняння драми Б. Шоу з газетним стилем

Групи фонем	Шоу		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	121,1	2992,71	130,1	6565,61	12,62	2,81	< 1%	Істотний
Передньо-язикові	390,2	9504,84	406,1	4991,76	15,54	4,10	< 0,1%	Істотний
Середньо-язикові	18,6	1557,56	7,3	1929,59	7,62	5,84	< 0,1%	Істотний
Задньо-язикові	68,5	2707,75	64,7	2702,39	9,50	1,57	> 10%	Випадковий
Носові	92,2	6098,84	82,3	1450,79	11,22	3,47	< 0,1%	Істотний
Сонорні	234,2	5611,44	225,9	4517,71	13,00	2,51	< 2%	Істотний
Щілинні	162,6	5351,56	183,4	1851,36	10,94	7,48	< 0,1%	Істотний
Зімкнені	202,6	5055,36	204,5	4025,75	12,30	0,6	> 50%	Випадковий

Таблиця А.56

Порівняння драми Б. Шоу з газетним стилем

Фонем на початку слова

Групи фонем	Шоу		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	59,1	3807,51	60,4	5795,16	12,65	0,40	> 50%	Випадковий
Передньо-язикові	95,8	1735,44	90,3	5333,09	10,85	2,00	5%	Випадковий
Середньо-язикові	16,4	1743,56	4,5	733	6,42	7,29	< 0,1%	Істотний
Задньо-язикові	33,6	1619,16	25,1	1089,91	6,72	5,00	< 0,1%	Істотний
Носові	11,5	441,75	5,3	384,39	3,71	6,58	< 0,1%	Істотний
Сонорні	71,4	3009,16	50,2	1738,84	8,90	9,38	< 0,1%	Істотний
Щілинні	71,6	3747,56	69,5	2357,75	10,08	0,82	> 20%	Випадковий
Зімкнені	62,2	1316,24	60,5	5259,75	10,46	0,64	> 50%	Випадковий

Таблиця А.57

Порівняння драми Б. Шоу з газетним стилем

Фонемі в кінці слова

Групи фонем	Шоу		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	26,2	742,84	20,5	563,75	4,66	4,82	< 0,1%	Істотний
Передньо-язикові	142,5	1247,75	134,9	5073,11	10,26	2,92	< 0,5%	Істотний
Середньо-язикові	–	–	–	–				
Задньо-язикові	15,4	633,36	14,6	473,56	4,30	0,73	> 20%	Випадковий
Носові	35,2	811,44	33	682	4,98	1,74	> 5%	Випадковий
Сонорні	60,7	1740,39	54,3	1752,39	7,62	3,31	< 0,2%	Істотний
Щілинні	48,3	1919,99	58,4	2055,36	8,14	4,40	< 0,1%	Істотний
Зімкнені	74,4	2069,16	56,0	3470	9,61	7,54	< 0,1%	Істотний

Таблиця А.58

Порівняння драми Б. Шоу з науковим стилем

Групи фонем	Шоу		Науковий стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	121,1	2992,71	123,5	5465,75	11,87	0,80	> 20%	Випадковий
Передньо-язикові	390,2	9504,84	425,1	11442,71	18,68	7,36	< 0,1%	Істотний
Середньо-язикові	18,6	1557,56	8,4	419,36	5,74	7,00	< 0,1%	Істотний
Задньо-язикові	68,5	2707,75	60,9	2282,71	9,12	3,28	< 0,2%	Істотний
Носові	92,2	6098,84	87,9	3125,91	12,40	1,37	> 10%	Випадковий
Щілинні	162,6	5351,56	186,8	2872,24	11,71	8,14	< 0,1%	Істотний
Зімкнені	202,6	5055,36	211,1	4202,71	12,42	2,69	< 1%	Істотний
Сонорні	234,2	5611,44	220,4	6461,16	14,18	3,83	< 0,1%	Істотний

Таблиця А.59

Порівняння драми Б. Шоу з науковим стилем

Фонемі на початку слова

Групи фонем	Шоу		Науковий стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	59,1	3807,51	52,0	2711,00	10,42	2,68	< 1%	Істотний
Передньо-язикові	95,8	1735,44	80,4	3005,36	8,89	6,82	< 0,1%	Істотний
Середньо-язикові	16,4	1743,56	1,5	57,75	5,48	10,71	< 0,1%	Істотний
Задньо-язикові	33,6	1619,16	21,0	1297,00	6,97	7,12	< 0,1%	Істотний
Носові	11,5	441,75	2,9	170,71	3,19	10,60	< 0,1%	Істотний
Сонорні	71,4	3009,16	30,4	687,16	7,85	20,57	< 0,1%	Істотний
Щілинні	71,6	3747,56	70,0	2138,00	9,90	0,64	> 50%	Випадковий
Зімкнені	62,2	1316,24	54,9	3164,71	8,64	3,33	< 0,2%	Істотний

Таблиця А.60

Порівняння драми Б. Шоу з науковим стилем

Фонемі в кінці слова

Групи фонем	Шоу		Науковий стиль		s	t	$2Q$	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	26,2	742,84	18,5	583,75	4,70	6,45	< 0,1%	Істотний
Передньо-язикові	142,5	1247,75	125,8	2447,44	7,85	8,38	< 0,1%	Істотний
Середньо-язикові	–	–	–	–				
Задньо-язикові	15,4	633,36	10,1	432,71	4,22	4,95	< 0,1%	Істотний
Носові	35,2	811,44	37,0	1889,00	6,71	1,06	> 20%	Випадковий
Сонорні	60,7	1740,39	54,2	3338,24	9,20	2,78	< 1%	Істотний
Щілинні	49,3	1919,99	56,5	1423,75	7,47	3,43	< 0,2%	Істотний
Зімкнені	74,4	2069,16	43,3	1220,79	7,40	16,54	< 0,1%	Істотний

Таблиця А.61

Порівняння наукового і газетного стилів

Групи фонем	Науковий стиль		Газетний стиль		S	t	$2Q$	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	123,5	5465,75	130,1	6565,51	14,16	1,84	5%	Випадковий
Передньо-язикові	425,1	11442,71	406,4	4991,76	16,55	4,44	< 0,1%	Істотний
Середньо-язикові	8,4	419,36	7,3	1929,59	6,26	0,69	> 20%	Випадковий
Задньо-язикові	60,9	2282,71	64,7	2702,39	9,12	1,64	> 10%	Випадковий
Носові	87,9	3125,91	82,3	1450,79	8,73	2,52	< 2%	Істотний
Сонорні	220,4	6461,16	225,9	4517,51	13,52	1,60	> 10%	Випадковий
Щілинні	186,8	2872,24	183,4	1851,36	8,87	1,51	> 10%	Випадковий
Зімкнені	211,1	4202,71	204,5	4025,75	11,71	2,22	< 5%	Істотний

Таблиця А.62

Порівняння наукового і газетного стилів

Фонеми на початку слова

Групи фонем	Науковий стиль		Газетний стиль		S	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	52,0	2711,00	60,4	5795,16	11,91	2,78	< 1%	Істотний
Передньо-язикові	80,4	3005,36	90,3	5333,09	11,78	3,31	< 0,2%	Істотний
Середньо-язикові	1,5	57,75	4,5	733	3,63	3,25	< 0,2%	Істотний
Задньо-язикові	21,0	1297,00	25,1	1089,91	6,31	2,56	< 2%	Істотний
Носові	2,9	170,71	5,3	384,39	3,04	3,11	< 0,5%	Істотний
Сонорні	30,4	687,16	50,2	1738,84	6,36	12,26	< 0,1%	Істотний
Щілинні	70,0	2138,00	69,5	2357,75	8,66	0,22	> 80%	Випадковий
Зімкнені	54,9	3164,71	60,5	5259,75	11,84	1,86	> 5%	Випадковий

Таблиця А.63

Порівняння наукового і газетного стилів

Фонемі в кінці слова

Групи фонем	Науковий стиль		Газетний стиль		s	t	2Q	Тип значення Величини $\bar{x}_1 - \bar{x}_2$
	\bar{x}	$\Sigma(x_i - \bar{x})^2$	\bar{x}	$\Sigma(x_i - \bar{x})^2$				
Губні	18,5	683,75	20,5	563,75	4,56	1,72	> 5%	Випадковий
Передньо-язикові	125,8	2447,44	134,9	5073,11	11,20	3,20	< 0,5%	Істотний
Середньо-язикові	–	–	–	–				
Задньо-язикові	10,1	432,71	14,6	473,56	3,88	4,56	< 0,1%	Істотний
Носові	37,0	1889,00	33	682	5,54	2,41	< 2%	Істотний
Сонорні	54,2	3338,24	54,3	1752,39	9,21	0,04	> 80%	Випадковий
Щілинні	56,5	1423,75	58,4	2055,36	7,61	0,98	> 20%	Випадковий
Зімкнені	43,3	1220,79	56,0	3470	8,84	5,66	< 0,1%	Істотний

Таблиця А.64

Результати зіставлення текстів за кількістю груп фонем

I – не врахування позиції фонему у слові; *II* – фонема на початку слова; *III* – фонема в кінці слова;

ПзБ – поезія Байрона; ПзМ – поезія Мура; Пр – художня проза; Др – драма; Гс – газетний стиль; Нс – науковий стиль; Рс – розмовний стиль

Стиль	ПзБ			ПзМ			Пр			Др			Гс			Нс			Рс		
	<i>I</i>	<i>II</i>	<i>III</i>	<i>I</i>	<i>II</i>	<i>III</i>	<i>I</i>	<i>II</i>	<i>III</i>	<i>I</i>	<i>II</i>	<i>III</i>	<i>I</i>	<i>II</i>	<i>III</i>	<i>I</i>	<i>II</i>	<i>III</i>	<i>I</i>	<i>II</i>	<i>III</i>
ПзБ	-	-	-	3	2	2	4	8	5	6	6	5	1	7	4	2	8	6	6	6	4
ПзМ	3	2	2	-	-	-	3	5	5	7	7	2	6	7	4	6	6	5	6	5	4
Пр	4	8	5	3	5	5	-	-	-	7	6	6	4	3	4	6	6	6	7	5	7
Др	6	6	5	7	7	2	7	6	6	-	-	-	6	4	5	6	7	6	4	4	4
Гс	1	7	4	6	7	4	4	3	4	6	4	5	-	-	-	3	6	4	6	6	3
Нс	2	8	6	6	6	5	6	6	6	6	7	6	3	6	4	-	-	-	7	7	4
Рс	6	6	4	6	5	4	7	5	7	4	4	4	6	6	3	7	7	4	-	-	-

Додатки Б.1 – Б.8. Результати отримані за критерієм Пірсона для 8 груп приголосних фонем

Таблиця Б.1

Дані групи губних фонем. Критерій Пірсона

№ інтервалів	$\alpha_i < X < \beta_i$	X_i	n_i	$(X_i - \bar{X})^2 n_i$	Z_i	$\varphi(z_i)$	n_i'	n_i' округлено	$\frac{(n_i - n_i')^2}{n_i}$
1	97,50–107,64	102,57	5	4832,94	-2,47	0,0184	1,85	2	0,05
2	107,64–117,79	112,72	6	2630,90	-1,67	0,0989	9,93	10	
3	117,79–127,93	122,86	23	2682,72	-0,86	0,2756	27,69	28	0,79
4	127,93–138,07	133,00	47	20,47	-0,05	0,3984	40,02	40	1,22
5	138,07–148,22	143,15	32	2881,92	0,76	0,2989	30,03	30	0,13
6	148,22–158,36	153,29	7	2697,36	1,57	0,1163	11,68	12	0,66
7	158,36–168,50	163,43	4	3545,01	2,38	0,0235	2,36	2	
Σ			124	19291,32			123,56	124	2,85

$$\bar{X} = 133,66; \quad S = \sqrt{\frac{19291,32}{123}} = 12,52; \quad \Delta X = 10,24$$

$$\frac{N\Delta X}{S} = 100,46; \quad \nu = 5-3=2; \quad \chi_{0,10;2}^2 = 4,71; \quad \chi^2 = 2,85; \quad P(\chi^2) > 10\%$$

Критерій Романовського

$$\frac{|\chi^2 - \nu|}{\sqrt{2\nu}} < 3; \quad \frac{|2,85 - 2|}{\sqrt{2 \cdot 2}} = 0,425 < 3$$

Таблиця Б.2

Дані групи передньоязикових фонем. Критерій Пірсона

№ інтервалів	$\alpha_i < X < \beta_i$	X_i	n_i	$(X_i - \bar{X})^2 n_i$	Z_i	$\varphi(z_i)$	n_i'	n_i' округлено	$\frac{(n_i - n_i')^2}{n_i}$
1	371–380	375,5	2	3732,48	-2,26	0,0310	1,81	2	0,04
2	380–389	384,5	5	5848,20	-1,79	0,0804	4,69		
3	389–398	393,5	15	9525,60	-1,32	0,1669	9,73	10	2,85
4	398–407	402,5	9	2361,96	-0,85	0,2780	16,22	16	3,21
5	407–416	411,5	22	1140,48	-0,38	0,3712	21,65	22	0,01
6	416–425	420,5	27	87,48	0,09	0,3972	23,18	23	0,63
7	425–434	429,5	18	2099,52	0,56	0,3410	19,89	20	0,18
8	434–443	438,5	12	4704,48	1,04	0,2323	13,55	14	0,18
9	443–452	447,5	10	8294,40	1,51	0,1276	7,44	7	0,38
10	452–461	456,5	2	2857,68	1,98	0,0562	3,27		
11	461–470	465,5	2	4380,48	2,45	0,0198	1,16		
Σ			124	45032,76			122,59	123	7,41

$$\bar{X} = 418,7; \quad s = \sqrt{\frac{45032,76}{123}} = 19,13; \quad \Delta X = 9$$

$$\frac{N\Delta X}{s} = 58,34; \quad \nu = 8-3=5; \quad \chi_{0,10;5}^2 = 9,24; \quad \chi^2 = 7,41; \quad P(\chi^2) > 10\%$$

Критерій Романовського

$$\frac{|\chi^2 - \nu|}{\sqrt{2\nu}} < 3; \quad \frac{|7,41 - 5|}{\sqrt{2 \cdot 5}} = 0,76 < 3$$

Таблиця Б.3

Дані групи середньоязикових фонем. Критерій Пірсона

№ інтервалів	$\alpha_i < X < \beta_i$	X_i	n_i	$(X_i - \bar{X})^2 n_i$	Z_i	$\varphi(z_i)$	n_i'	n_i' округ-лено	$\frac{(n_i - n_i')^2}{n_i}$
1	0,5–1,5	1	2 } 6 } 8	68,68	-1,94	0,0608	2,50 } 4,50 } 7	3	0,14
2	1,5–2,5	2		141,72	-1,61	0,1092		5	
3	2,5–3,5	3	8	119,20	-1,28	0,1758	7,22	7	0,08
4	3,5–4,5	4	13	106,33	-0,95	0,2541	10,43	10	0,63
5	4,5–5,5	5	20	69,19	-0,62	0,3292	13,52	14	3,11
6	5,5–6,5	6	10	7,40	-0,28	0,3836	15,75	16	2,10
7	6,5–7,5	7	12	0,24	0,05	0,3984	16,36	16	1,16
8	7,5–8,5	8	12	15,60	0,38	0,3712	15,24	15	0,69
9	8,5–9,5	9	15	68,69	0,71	0,3101	12,72	13	0,41
10	9,5–10,5	10	12	118,32	1,04	0,2323	9,54	10	0,63
11	10,5–11,5	11	6	102,84	1,37	0,1561	6,41	6	0,03
12	11,5–12,5	12	2 } 1 } 4 } 7	52,84	1,70	0,0940	3,86 } 2,08 } 1,00 } 6,94	4	0,00
13	12,5–13,5	13		37,70	2,03	0,0508		2	
14	13,5–14,5	14		203,92	2,36	0,0246		1	
Σ			123	1112,65			121,13	122	8,98

$$\bar{X} = 6,86; \quad S = \sqrt{\frac{1112,65}{122}} = 3,02; \quad \Delta X = 1 \quad \frac{N\Delta X}{S} = 40,73; \quad \nu = 11 - 3 = 8; \quad \chi_{0,25;8}^2 = 10,2; \quad \chi^2 = 8,98; \quad P(\chi^2) > 25\%$$

Критерій Романовського

$$\frac{|\chi^2 - \nu|}{\sqrt{2\nu}} < 3; \quad \frac{|8,98 - 8|}{\sqrt{2 \cdot 8}} = 0,245 < 3$$

Таблиця Б.4

Дані групи задньоязикових фонем. Критерій Пірсона

№ інтервалів	$\alpha_i < X < \beta_i$	X_i	n_i	$(X_i - \bar{X})^2 n_i$	Z_i	$\varphi(z_i)$	n_i'	n_i' округ-лено	$\frac{(n_i - n_i')^2}{n_i}$		
1	30–34	32,0	1 } 0 } 3 } 1 }	850,31	-3,10	0,0033	0,19 } 0,58 } 1,64 } 3,96 }	0	0,29		
2	34–38	36,0		0	-2,68	0,0110		0,58		1	
3	38–42	40,0		3	1343,87	-2,26		0,0310		1,64	2
4	42–46	44,0		1	294,64	-1,83		0,0748		3,96	4
5	46–50	48,0	8	1385,54	-1,40	0,1497	7,92	8	0,01		
6	50–54	52,0	16	1343,96	-0,98	0,2468	13,06	13	0,66		
7	54–58	56,0	15	400,16	-0,55	0,3429	18,14	18	0,54		
8	58–62	60,0	23	31,22	-0,12	0,3961	20,95	21	0,20		
9	62–66	64,0	18	148,78	0,30	0,3814	20,18	20	0,24		
10	66–70	68,0	17	794,19	0,73	0,3056	16,17	16	0,04		
11	70–74	72,0	11	1291,37	1,16	0,2036	10,77	11	0,00		
12	74–78	76,0	7 } 4 }	1540,54	1,58	0,1145	6,06 } 2,80 }	6	0,52		
13	78–82	80,0		4	1419,03	2,01		0,0529		2,80	3
Σ			124	10844,61			122,42	123	2,50		

$$\bar{X} = 61,16; \quad S = \sqrt{\frac{10844,61}{123}} = 9,39; \quad \Delta X = 4 \quad \frac{N\Delta X}{S} = 52,82; \quad \nu = 9-3=6; \quad \chi_{0,75;6}^2 = 3,45; \quad \chi^2 = 2,50; \quad P(\chi^2) > 75\%$$

Критерій Романовського

$$\frac{|\chi^2 - \nu|}{\sqrt{2\nu}} < 3; \quad \frac{|2,50 - 6|}{\sqrt{2 \cdot 6}} = 1,01 < 3$$

Таблиця Б.5

Дані групи носових фонем. Критерій Пірсона

№ інтервалів	$\alpha_i < X < \beta_i$	X_i	n_i	$(X_i - \bar{X})^2 n_i$	Z_i	$\varphi(z_i)$	n_i	n_i округлено	$\frac{(n_i - n_i')^2}{n_i}$
1	62,50–68,64	65,57	10	3175,52	-1,92	0,0632	5,14	5	4,59
2	68,64–74,78	71,71	14	1909,91	-1,27	0,1781	14,50	15	0,02
3	74,78–80,92	77,85	19	583,14	-0,60	0,3332	27,00	27	2,37
4	80,92–87,06	83,99	35	12,60	-0,06	0,3982	32,41	32	0,21
5	87,06–93,20	90,13	30	1362,83	0,73	0,3056	24,90	25	1,04
6	93,20–99,34	96,27	10	1658,94	1,40	0,1497	12,19	12	0,39
7	99,34–105,5	102,4	50	1808,80	2,05	0,0488	3,97	4	0,27
Σ			123	10511,75			120,11	120	8,89

$$\bar{X} = 83,39; \quad s = \sqrt{\frac{10511,75}{122}} = 9,28; \quad \Delta X = 6,14$$

$$\frac{N\Delta X}{s} = 81,38; \quad \nu = 7-3=4; \quad \chi_{0,05;4}^2 = 9,49; \quad \chi^2 = 8,89; \quad P(\chi^2) > 5\%$$

Критерій Романовського

$$\frac{|\chi^2 - \nu|}{\sqrt{2\nu}} < 3; \quad \frac{|8,89 - 4|}{\sqrt{2 \cdot 4}} = 1,73 < 3$$

Таблиця Б.6

Дані групи сонорних фонем. Критерій Пірсона

№ інтервалів	$\alpha_i < X < \beta_i$	X_i	n_i	$(X_i - \bar{X})^2 n_i$	Z_i	$\varphi(z_i)$	n_i'	n_i' округлено	$\frac{(n_i - n_i')^2}{n_i}$
1	189,50–195,41	192,455	3 } 11 } 14	1766,10	-1,89	0,0681	3,08 } 5,68 } 8,76	3	3,13
2	195,41–201,32	198,365		6578,52	-1,52	0,1257		6	
3	201,32–207,23	204,275	12	4127,00	-1,15	0,2059	9,30	9	0,78
4	207,23–213,14	210,185	12	1915,72	-0,78	0,2943	13,30	13	0,13
5	213,14–219,05	216,095	14	633,16	-0,42	0,3653	16,50	17	0,38
6	219,05–224,96	222,005	14	9,30	-0,05	0,3984	18,00	18	0,89
7	224,96–230,87	227,915	13	337,47	0,32	0,3790	17,12	17	0,99
8	230,87–236,78	233,825	16	1937,76	0,68	0,3166	14,30	14	0,20
9	236,78–242,69	239,735	12	3433,41	1,05	0,2299	10,39	10	0,25
10	242,69–248,60	245,645	11	5730,79	1,42	0,1456	6,58	7	2,97
11	248,60–254,51	251,555	5	4128,50	1,79	0,0804	3,63	4	0,52
Σ			123	31597,72			117,88	118	10,24

$$\bar{X} = 222,82; \quad S = \sqrt{\frac{31597,72}{122}} = 16,09; \quad \Delta X = 5,91$$

$$\frac{N\Delta X}{S} = 45,18; \quad \nu = 10 - 3 = 7; \quad \chi_{0,10;7}^2 = 12,02; \quad \chi^2 = 10,24; \quad P(\chi^2) > 10\%$$

Критерій Романовського

$$\frac{|\chi^2 - \nu|}{\sqrt{2\nu}} < 3; \quad \frac{|10,24 - 7|}{\sqrt{2 \cdot 7}} = 0,87 < 3$$

Таблиця Б.7

Дані групи щільних фонем. Критерій Пірсона

№ інтервалів	$\alpha_i < X < \beta_i$	X_i	n_i	$(X_i - \bar{X})^2 n_i$	Z_i	$\varphi(z_i)$	n_i'	n_i' округ-лено	$\frac{(n_i - n_i')^2}{n_i}$
1	156,5–165,0	160,77	3	4419,99	-2,06	0,0478	2,72	3	0,45
2	165,0–173,6	169,32	4	3561,22	-1,60	0,1109	6,30	6	
3	173,6–182,1	177,86	12	5441,21	-1,14	0,2083	11,84	12	0,00
4	182,1–190,7	186,41	26	4225,30	-0,68	0,3166	18,00	18	3,55
5	190,7–199,2	194,95	25	441,84	-0,23	0,3894	22,14	22	0,37
6	199,2–207,8	203,50	17	320,50	0,23	0,3885	22,09	22	1,17
7	207,8–216,3	212,04	16	2656,78	0,69	0,3144	17,87	18	0,20
8	216,3–224,9	220,59	8	3674,64	1,15	0,2059	11,70	12	1,17
9	224,9–233,4	229,13	6	5391,36	1,61	0,1092	6,21	6	0,01
10	233,4–241,9	237,68	3	5935,78	2,07	0,0468	2,66	3	3,24
11	241,9–250,5	246,22	4	6645,62	2,53	0,0163	0,93	1	
Σ			124	42714,24			122,46	123	10,16

$$\bar{X} = 199,156; \quad S = \sqrt{\frac{42714,24}{123}} = 18,635; \quad \Delta X = 8,545$$

$$\frac{N\Delta X}{S} = 56,86 \quad \nu = 9-3=6; \quad \chi_{0,10,7}^2 = 10,65; \quad \chi^2 = 10,16; \quad P(\chi^2) > 10\%$$

Критерій Романовського

$$\frac{|\chi^2 - \nu|}{\sqrt{2\nu}} < 3; \quad \frac{|10,16 - 6|}{\sqrt{2 \cdot 6}} = 1,20 < 3$$

Таблиця Б.8

Дані групи зімкнених фонем. Критерій Пірсона

№ інтервалів	$\alpha_i < X < \beta_i$	X_i	n_i	$(X_i - \bar{X})^2 n_i$	Z_i	$\varphi(z_i)$	n_i'	n_i' округлено	$\frac{(n_i - n_i')^2}{n_i}$
1	150,50–158,14	154,32	5 } 4 } ₉	9585,19	-2,23	0,0332	1,60	2	2,83
2	158,14–165,77	161,95		5226,71	-1,84	0,0734	3,55		
3	165,77–173,41	169,59	8	6503,47	-1,45	0,1394	6,74	7	0,23
4	173,41–181,04	177,23	9	3922,27	-1,06	0,2275	10,99	11	0,36
5	181,04–188,68	184,86	9	1577,68	-0,68	0,3166	15,29	15	2,58
6	188,68–196,32	192,50	13	408,26	-0,29	0,3825	18,48	18	1,62
7	196,32–203,95	200,13	25	103,23	0,10	0,3970	19,18	19	1,76
8	203,95–211,59	207,77	19	1775,93	0,49	0,3538	17,09	17	0,19
9	211,59–219,22	215,41	15	4491,43	0,88	0,2709	13,08	13	0,28
10	219,22–226,86	223,04	10	6220,04	1,27	0,1781	8,61	9	0,22
11	226,86–234,50	230,68	7	7428,37	1,66	0,1006	4,86	5	
Σ			124	47242,58			119,47	120	0,94

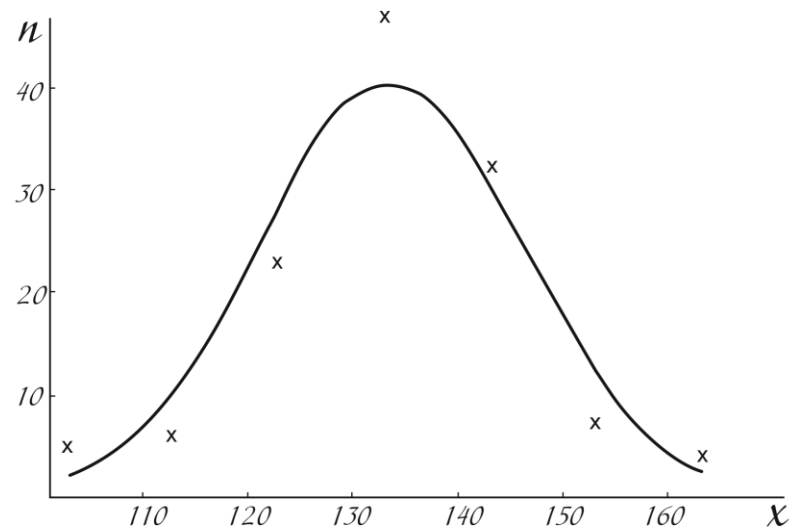
$$\bar{X} = 198,102; \quad S = \sqrt{\frac{47242,58}{123}} = 19,60; \quad \Delta X = 7,636$$

$$\frac{N\Delta X}{S} = 48,31 \quad \nu = 10 - 3 = 7; \quad \chi_{0,10;7}^2 = 12,02; \quad \chi^2 = 11,06; \quad P(\chi^2) > 10\%$$

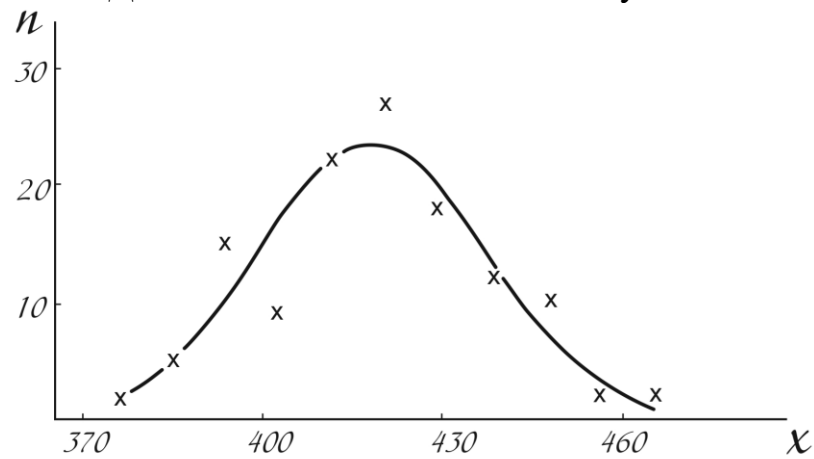
Критерій Романовського

$$\frac{|\chi^2 - \nu|}{\sqrt{2\nu}} < 3; \quad \frac{|11,06 - 7|}{\sqrt{2 \cdot 7}} = 1,08 < 3.$$

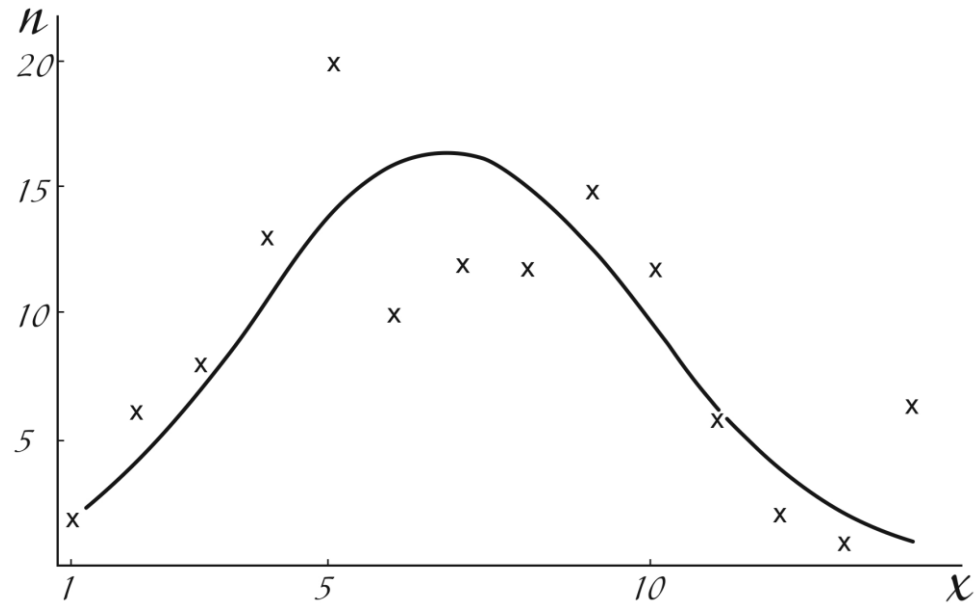
Додатки В.1 – В.8. Графіки нормального розподілу імовірностей частот 8 груп приголосних фонем



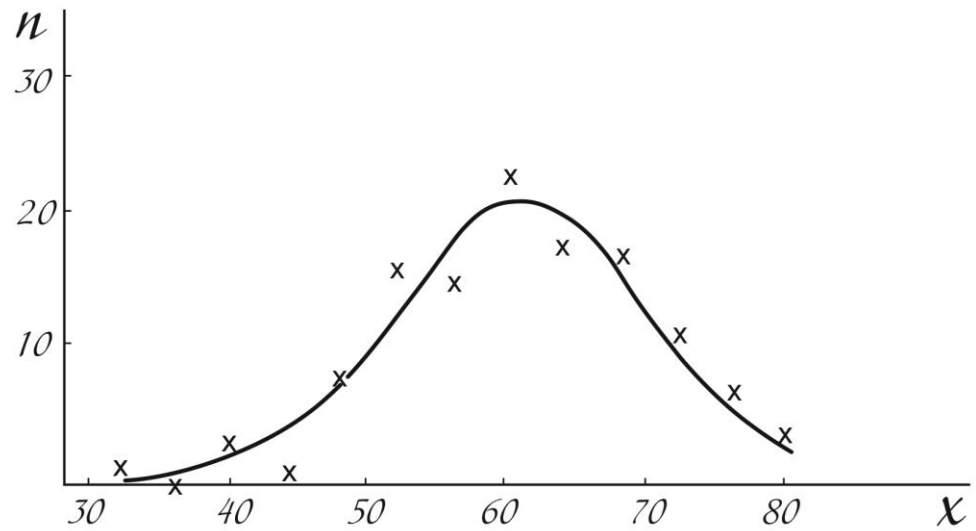
Додаток В.1. Розподіл частоти губних



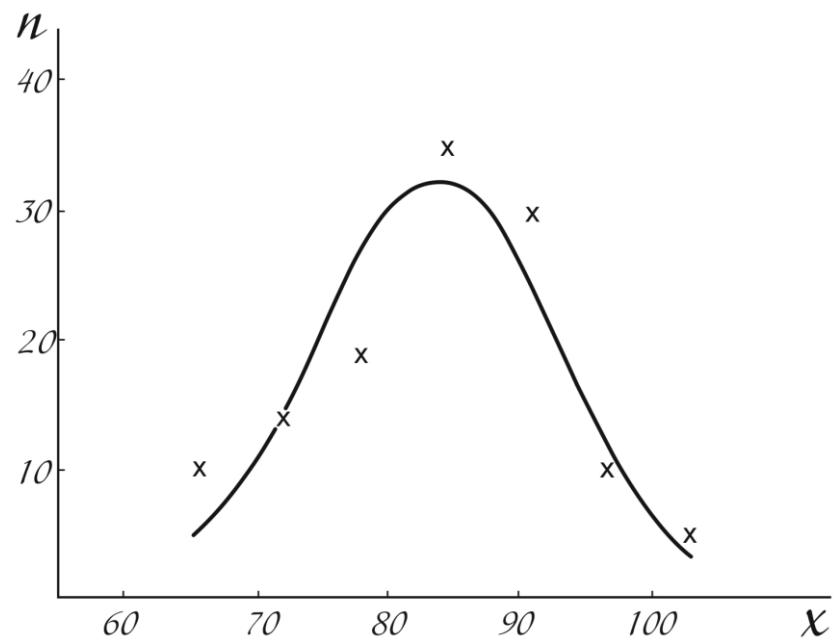
Додаток В.2. Розподіл частоти передньоязикових



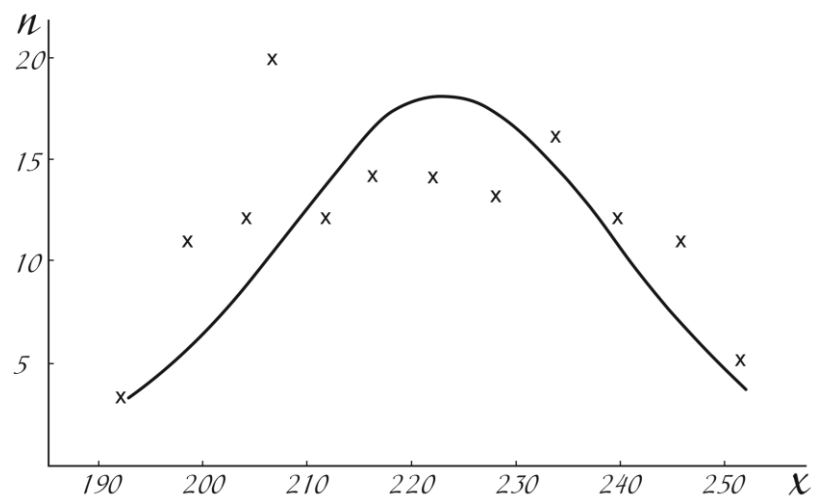
Додаток В.3. Розподіл частоти середньоязикових



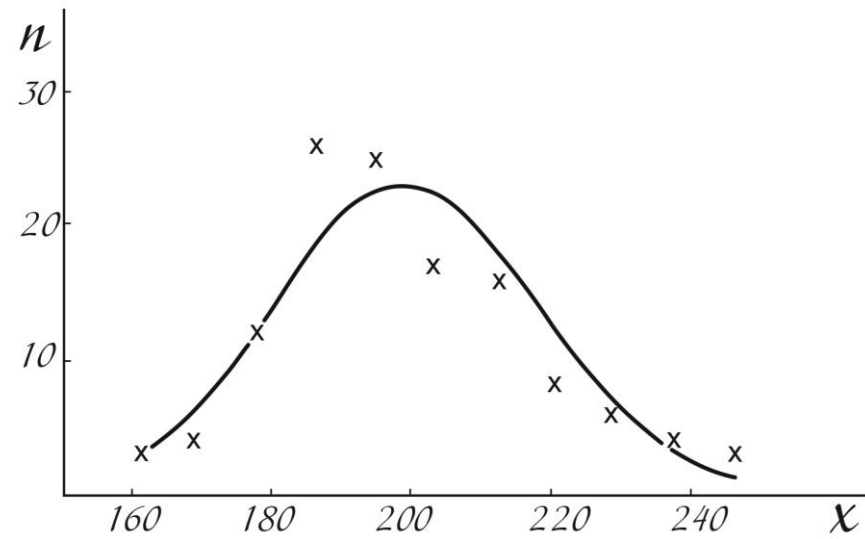
Додаток В.4. Розподіл частоти задньоязикових



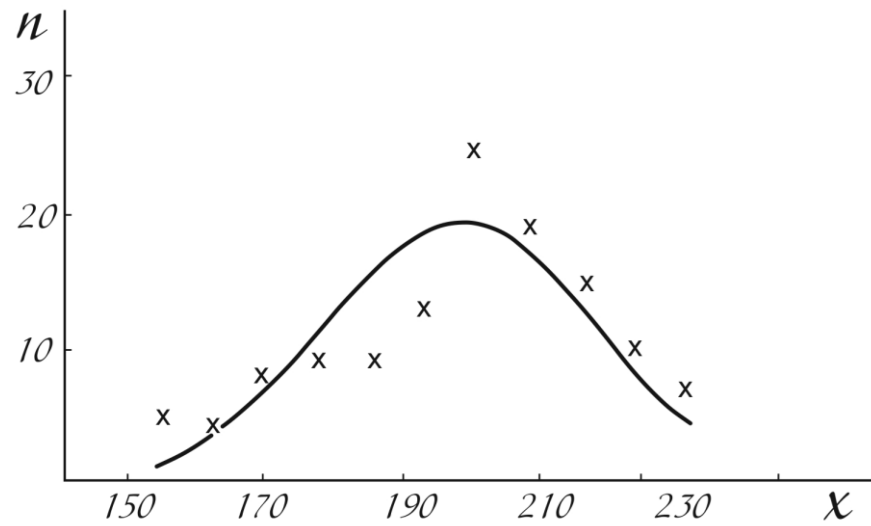
Додаток В.5. Розподіл частоти носових



Додаток В.6. Розподіл частоти сонорних

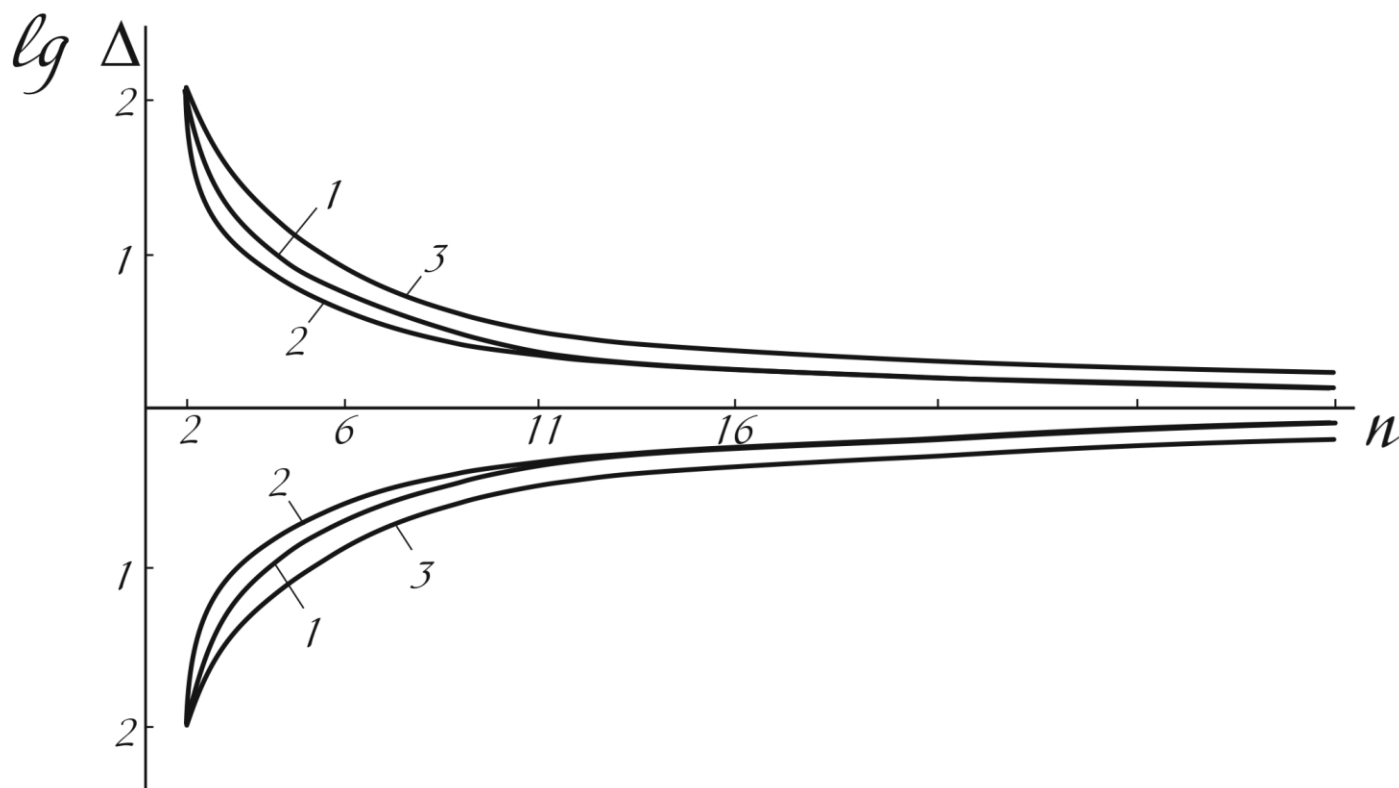


Додаток В.7. Розподіл частоти щільних



Додаток В.8. Розподіл частоти зімкнених

Додаток Д. Графік залежності величини Δ від кількості порцій вибірки n для групи середньоязикових



- 1/ художня проза Байрона;
- 2/ поезія Байрона;
- 3/ науковий стиль.

Додатки Е.1 – Е.6. Результати за методами ранжування та визначення відстаней між стилями

Таблиця Е.1

Результати за методом ранжування. Неврахування позиції фонему у слові

Стилі Групи фонем	ПзБ	ПзМ	Пр	Др	Гс	Нс	Рс
Губні	5 134,3 3	5 137,9 2	5 139,7 1	5 121,1 7	5 130,1 5	5 123,5 6	5 131,9 4
Передньо-язикові	1 421,8 3	1 425,0 2	1 402,7 5	1 390,2 6	1 406,4 4	1 425,1 1	1 362,9 7
Середньо-язикові	8 6,9 5	8 5,9 7	8 6,7 6	8 18,6 1; 2	8 7,3 4	8 8,4 3	7 18,6 1; 2
Задньо-язикові	7 64,3 3	7 59,3 6	7 60,9 4; 5	7 68,5 1	7 64,7 2	7 60,9 4; 5	8 12,6 7
Носові	6 83,6 3	6 82,9 4	6 80,0 6	6 92,2 1	6 82,3 5	6 87,9 2	6 76,8 7
Сонорні	2 224,7 5	2 233,9 2	2 210,4 7	2 234,2 1	2 225,9 4	2 220,4 6	2 226,9 3
Щілинні	4 190,3 3	3 210,3 1	3 204,2 2	4 162,6 6	4 183,4 5	4 186,8 4	4 158,9 7
Зімкнені	3 202,9 4	4 182,7 7	4 194,1 6	3 202,6 5	3 204,5 3	3 211,1 2	3 226,5 1

Таблиця Е.2

Результати за методом ранжування. Фонема на початку слова

Стилі / Групи фонем	ПзБ	ПзМ	Пр	Др	Гс	Нс	Рс
Губні	3 76,1 1	3 71,8 3	3 64,0 4	5 59,1 6	4 60,4 5	4 52,0 7	3 72,0 2
Передньо-язикові	1 111,3 2	1 115,7 1	1 93,2 5	1 95,8 4	1 90,3 6	1 80,4 7	1 100,0 3
Середньо-язикові	8 2,7 4	8 2,3 5	8 1,3 7	7 16,4 1	8 4,5 3	8 1,5 6	7 11,6 2
Задньо-язикові	6 36,8 1	6 31,8 3	6 30,6 5	6 33,6 2	6 25,1 6	6 21,0 7	6 31,3 4
Носові	7 9,8 2	7 7,7 3	7 6,7 4	8 11,5 1	7 5,3 6	7 2,9 7	8 5,7 5
Сонорні	5 63,5 3	4 63,1 4	5 48,4 6	3 71,4 1	5 50,2 5	5 30,4 7	5 68,4 2
Щілинні	2 97,6 2	2 101,9 1	2 83,9 3	2 71,6 4	2 69,5 6	2 70,0 5	4 69,3 7
Зімкнені	4 66,8 2	5 57,1 5	4 56,8 6	4 62,2 3	3 60,5 4	3 54,9 7	2 78,2 1

Таблиця Е.3

Результати за методом ранжування. Фонема в кінці слова

Стилі Групи фонем	ПзБ	ПзМ	Пр	Др	Гс	Нс	Рс
Губні	6 22,1 5	6 26,4 2	6 29,2 1	6 26,2 3	6 20,5 6	6 18,5 7	6 23,5 4
Передньо-язикові	1 165,1 1	1 161,1 2	1 135,2 4	1 142,5 3	1 134,9 5	1 125,8 7	1 128,7 6
Середньо-язикові	–	–	–	–	–	–	–
Задньо-язикові	7 12,4 5	7 14,8 4	7 8,6 7	7 15,4 2	7 14,9 3	7 10,1 8	7 18,4 1
Носові	5 33,0 5; 6	5 34,4 3; 4	5 30,7 7	5 35,2 2	5 33,0 5; 6	5 37,0 1	5 34,4 3; 4
Сонорні	4 55,3 3	4 57,0 2	4 48,0 7	3 60,0 1	4 54,3 4	3 54,2 5	3 53,6 6
Щілинні	3 66,1 2	3 71,0 1	2 65,0 3	4 48,3 7	2 58,4 4	2 56,5 5	4 49,8 6
Зімкнені	2 79,0 1	2 74,5 2	3 58,1 5	2 74,4 3	3 56,0 6	4 43,3 7	2 67,6 4

Таблиця Е.4

Результати за методом визначення відстаней між стилями

Відстані стилів від розмовного стилю

Позицію фонем у слові не враховано

Групи фонем	Поезія Дж. Г. Байрона	Художня проза	Поезія Т. Мура	Драма	Науковий стиль	Газетний стиль
Губні	–	0,12	–	0,38	0,11	–
Передньо-язикові	0,78	0,67	0,79	0,51	0,78	0,71
Середньо-язикові	0,58	0,59	0,62	–	0,52	0,51
Задньо-язикові	0,29	0,52	0,55	–	0,52	0,26
Носові	0,15	–	0,16	0,57	0,49	0,10
Сонорні	–	0,44	–	–	–	–
Щілинні	0,71	0,80	0,84	–	0,78	0,73
Зімкнені	0,70	0,72	0,81	0,72	0,57	0,70

Таблиця Е.5

Результати за методом визначення відстаней між стилями

Відстані стилів від розмовного стилю

Фонема на початку слова

Групи фонем	Поезія Дж. Г. Байрона	Художня проза	Поезія Т. Мура	Драма	Науковий стиль	Газетний стиль
Губні	–	0,36	–	0,52	0,71	0,41
Передньо-язикові	0,41	–	0,55	–	0,65	0,22
Середньо-язикові	0,56	0,63	0,58	0,02	0,63	0,40
Задньо-язикові	0,11	–	–	–	0,59	0,33
Носові	0,48	–	0,14	0,68	0,47	–
Сонорні	–	0,70	–	–	0,87	0,70
Щілинні	0,79	0,47	0,80	–	–	–
Зімкнені	0,49	0,74	0,73	0,66	0,74	0,62-

Таблиця Е.6

Результати за методом визначення відстаней між стилями

Відстані стилів від розмовного стилю

Фонема в кінці слова

Групи фонем	Поезія Дж. Г. Байрона	Художня проза	Поезія Т. Мура	Драма	Науковий стиль	Газетний стиль
Губні	–	0,34	–	–	0,33	–
Передньо-язикові	0,81	0,02	0,81	0,64	–	–
Середньо-язикові						
Задньо-язикові*	0,58	0,75	0,20	0,12	0,70	0,34
Носові	–	0,04	–	–	–	–
Сонорні	–	0,32	–	0,47	–	–
Щілинні	0,71	0,58	0,75	–	0,34	0,45
Зімкнені	0,45	0,53	0,29	0,68	0,84	0,57

* Середньоязикових у кінці слова нема

Додатки Ж.1 – Ж.7. Процес, цикл формування вибірки. Процес обчислення

```

private String prepareConsonants(final String transcription, final String
fileName) throws WrongDataException {
    final String onlyAlphabetic = transcription.replaceAll("[^a-zA-Z0
-9_]", "");
    final String onlyConsonant = transcription.replaceAll("[
AEIOUYaeiouy]", "");
    if(onlyConsonant.length() < globalLength) {
        throw new WrongDataException("Not enough consonants in text:
" + fileName);
    }

    return onlyConsonant.substring(0, globalLength);
}

```

Рис. Ж.1. Процес формування вибірки

```

private List<String> separateTextIntoGroups(final String text) {
    final List<String> groups = Lists.newArrayList();
    final char [] chars = text.toCharArray();

    for(int i = 0; i < chars.length; i = i + groupLength) {
        if(i + groupLength <= chars.length) {
            addNewElement(groups, chars, i, groupLength);
            continue;
        }
        addNewElement(groups, chars, i, chars.length - i);
    }
    return groups;
}

private void addNewElement(final List<String> groups,
final char [] chars,
final int startIndex,
final int endIndex) {
    final char [] temp = new char [endIndex];
    System.arraycopy(chars, startIndex, temp, 0, endIndex);
    groups.add(new String(temp));
}

```

Рис. Ж.2. Цикл формування порцій

```

public MultiValueMap<ConsonantType, Long> processConsonants(final String
transcription, final String fileName)
throws WrongDataException {
    final String consonants = prepareConsonants(transcription,
fileName);
    final List<String> consonantGroups = separateTextIntoGroups(
consonants);

    return createMultiValueMap(consonantGroups);
}

private MultiValueMap<ConsonantType, Long> createMultiValueMap(List<
String> consonantGroups) {

```

Рис. Ж.3. Процес обчислення для кожної групи фонем

```

public boolean pearsonTest(final List<Long> sample) {
    final long min = sample .stream().mapToLong(x->x).min().
        getAsLong();
    final long max = sample .stream().mapToLong(x->x).max().
        getAsLong();

    final double intervalLength = (max - min) / INTERVAL_COUNT; //
        findIntervalLength(min, max);
    final Map<Double, Long> sets = splitToSets(sample, min, max,
        intervalLength);

    final double sampleMean = mean(sets); //  $\mu$ 
    final double standardDeviation = Math.sqrt(variance(sampleMean,
        sets)); //  $\sigma$ 

```

Рис. Ж.4. Процес обчислення за критерієм Пірсона

```

public class StudentTest implements StatisticTest {

    @Value ("${student.t.value}")
    private final static double T;

    public boolean test(final List<Long> sample1, final List<Long>
        sample2) {
        double mean1 = mean(sample1);
        double mean2 = mean(sample2);
        double variance1 = variance(mean1, sample1);
        double variance2 = variance(mean2, sample2);
        double tValue = test(mean1, mean2, variance1, variance2, sample1.
            size(), sample2.size());
        return Math.abs(tValue) <= T;
    }

    private double test(final double mean1, final double mean2,

        final double variance1, final double variance2,
        final double n1, final double n2) {
        return (mean1 - mean2) / Math.sqrt((n1 - 1) * variance1 + (n2 - 1) *
            variance2)
            * Math.sqrt((n1 * n2 * (n1 + n2 - 2)) / (n1 + n2));
    }

    private double mean(List<Long> sample) {
        return 1.0 / sample.size() * sample.stream()
            .mapToLong(x->x)
            .sum();
    }

    private double variance(double mean, List<Long> sample) {
        return 1.0 / (sample.size() - 1) * sample.stream()
            .mapToDouble(x->Math.pow(x - mean, 2))
            .sum();
    }

```

Рис. Ж.5. Процес обчислення за критерієм Стьюдента

```

public boolean test(List<Long> sample1 ,List<Long> sample2 ) throws
WrongDataException {
    final double sum1 = calculateSumOfList(sample1 );
    final double sum2 = calculateSumOfList(sample2 );
    final List<Double> empiricalFrequency1 = countEmpiricalFrequency(
        sample1 , sum1 );
    final List<Double> empiricalFrequency2 = countEmpiricalFrequency(
        sample2 , sum2 );
    final List<Double> accumulatedEmpiricalFrequency1 =
        countAccumulatedEmpiricalFrequency(empiricalFrequency1);
    final List<Double> accumulatedEmpiricalFrequency2 =
        countAccumulatedEmpiricalFrequency(empiricalFrequency2);

    final List<Double> absDiff=
        calculateAbsDiff(accumulatedEmpiricalFrequency1 ,
            accumulatedEmpiricalFrequency2);
    final double dMax= findTheLargestAbsoluteDifference(absDiff);
    final double empiricalValueOfTheKolmogorovCriterion =
        calculateEmpiricalValueOfTheKolmogorovCriterion(dMax,
            sum1, sum2 );
    return empiricalValueOfTheKolmogorovCriterion<LAMBDA;
}

```

Рис. Ж.6. Процес обчислення за критерієм Колмогорова-Смірнова

```

// ( s1 , s2 , s1 + s2 )
private List<Triple<Long , Long , Long>> getSumOfTwoSamples ( final List
<Long
    > sample1 , final List<Long> sample2 ) {
    final List<Triple<Long , Long , Long>> result = new ArrayList<>();
    for ( int i = 0 ; i < sample1 . size () ; i ++ ) {
        result.add ( Triple . of ( sample1 . get ( i ) , sample2 . get ( i ) ,
            sample1 . get ( i ) + sample2 . get ( i ) ) ) ;
    }
}

```

Рис. Ж.7. Процес обчислення за критерієм хі-квадрат

Додаток К

Акт впровадження результатів дисертаційного дослідження



АКТ

про впровадження в навчальний процес кафедри прикладної лінгвістики
Національного університету "Львівська політехніка" результатів
дисертаційної роботи Хомицької Ірини Юріївни
на тему «Методи та засоби диференціації фоностатистичних структур
функціональних стилів англійської мови»

Даний акт складений про те, що результати дисертаційної роботи Хомицької Ірини Юріївни впроваджені і використані в навчальному процесі кафедри прикладної лінгвістики Національного університету "Львівська політехніка" для студентів спеціальності 035 «Філологія». Зокрема при викладанні дисциплін:

- «Основи прикладної лінгвістики» – метод комплексного аналізу диференціації фоностатистичних структур стилів та багатофакторний метод визначення ступеню дії факторів стилю, підстилю та авторської манери викладу. Методи дають змогу ознайомити студентів з сучасним математичним забезпеченням технологій авторської атрибуції тексту.

- «Контрастивна стилістика англійської та української мов» – статистичні моделі стилевої, підстилевої та авторської диференціації за методом гіпотез та методом ранжування і статистичну модель визначення стилерозрізняльної здатності груп приголосних фонем. Моделі дають змогу ознайомити студентів з сучасним математичним забезпеченням стилевої, підстилевої та авторської диференціації текстів.

Доцент кафедри ПЛ,
к.ф.н., доцент

Н. І. Романишин

Завідувач кафедри ПЛ,
д.ф.н., професор

О. П. Левченко

Додаток Л

Список публікацій здобувача 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*. July 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., Holovatyy 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, 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.

Додаток М

Апробація роботи

Основні теоретичні положення та практичні результати дисертаційної роботи доповідалися і обговорювались на: 11-й Міжнародній науково-технічній конференції Інституту комп'ютерних наук та інформаційних технологій, Національного університету “Львівська політехніка” 2016; 14-й науково-технічній Міжнародній конференції "Досвід проектування та застосування систем автоматизованого проектування в мікроелектроніці" кафедри систем автоматизованого проектування, Національного університету “Львівська політехніка” 2017; 12-й Міжнародній науково-технічній конференції Інституту комп'ютерних наук та інформаційних технологій, Національного університету “Львівська політехніка” 2017; 13-й Міжнародній науково-технічній конференції Інституту комп'ютерних наук та інформаційних технологій, Національного університету “Львівська політехніка” 2018; 15-й Міжнародній науково-технічній конференції "Досвід проектування та застосування систем автоматизованого проектування" кафедри систем автоматизованого проектування, Національного університету “Львівська політехніка” 2019; 14-й Міжнародній науково-технічній конференції Інституту комп'ютерних наук та інформаційних технологій, Національного університету “Львівська політехніка” 2019; 15-й Міжнародній науково-технічній конференції Інституту комп'ютерних наук та інформаційних технологій, Національного університету “Львівська політехніка” 2020; 4-й Міжнародній науково-технічній конференції "Комп'ютерна лінгвістика та інтелектуальні системи" 2020; 14-й науково-практичній конференції “Проблеми та перспективи розвитку економіки і підприємництва та комп'ютерних технологій в Україні” Інституту підприємництва та перспективних технологій, Національного університету “Львівська політехніка” 2018.