обрядово-ігровим текстом за допомогою сучасних засобів інформації, адже «відмінність між людьми в природі та якості їх образів відображають відмінність інформації, до збору якої вони приготувались» [Найссер 1981, с. 144]. В такому випадку взаємодія фреймів двох текстів, джерела та реципієнта, буде здійснюватись за наступною схемою: образ було сформовано шляхом стимуляції всіх сенсорних систем, отже і вхід до образу буде відбуватись не через окрему грань, а через всі грані одночасно. Наприклад, назва статті "***A red rag** to a **bull**ish taxman, surely?*" (Private Eye № 1213) відсилає адресата до обрядово-ігрового тексту «Корида». Актуалізація змісту «річ, яка дратує податківця» відбувається таким шляхом: ідентифікація елементів верхнього рівня фрейму «Корида» (***A red rag,*** корінь ***bull-*** одиниці ***bull**ish*) > відтворення фрейму «Корида» повністю > взаємодія фрейму «Корида» з фреймом статті.

Отже, актуалізація змісту інтермедіальних алюзій шляхом взаємодії фреймів тексту-джерела та тексту-реципієнта відбувається наступним чином:

1) віднаходження еквівалента вербальному знаку серед аудіо / візуальних / одористичних / смакових / тактильних знаків (АБО серед аудіо і візуальних знаків, якщо текст-джерело створений в межах семіотичної системи, що складається зі знаків, які апелюють до двох відчуттів – зору та слуху);

2) побудова відповідної (аудіо / візуальної / одористичної / смакової / тактильної) грані фрейму (АБО двох – аудіо і візуальної – граней);

3) відтворення всіх інших граней фрейму;

4) наповнення фрейму інформацією з тексту, в якому вжито інтермедіальні алюзивні засоби.

Перспективним напрямком дослідження вважаємо детальне вивчення особливостей відтворення різних типів (предметного, акціонального, посесивного, таксономічного і компаративного – за С. А. Жаботинською [Жаботинская 1999]) фреймів текстів за вербалізованою складовою у процесі актуалізації змісту інтермедіальної алюзії.

Література

Басин Е. Искусство // Энциклопедия «Кругосвет» // http://www.krugosvet.ru/articles/104/1010471/1010471a2.htm (Cited 18/12/2008) *Бахтин М. М.* Эстетика словесного творчества/ Сост. С. Г. Бочаров. – 2-е изд. – М.: Искусство, 1986. – 445с. *Бондалетова О. Б.* Представление знаний: фреймы и понятия // Философские науки. – 1989. – № 4. – С. 37-43. *Жаботинская С. А.* Концептуальный анализ: типы фреймов // Вісник Черкаського університету. Серія Філологічні науки. – Черкаси, 1999. – Вип.11. – С. 12-25. *Еко У.* Роль читача. Дослідження з семіотики текстів / Пер. з англійської Мар'яни Гірняк. – Львів: Літопис, 2004. – 384 с. *Моррис Ч.У.* Основания теории знаков // Семиотика. Сборник переводов. Под ред. Ю. С. Степанова. – М.: Радуга, 1982. *Найссер У.* Познание и реальность: Смысл и принципы когнитивной психологии. – М.: «Прогресс», 1981. – 232 с. *McConnell, James V.* Understanding Human Behavior. Sixth Edition. – The University of Michigan, 1988. – 646 p.

ВОЛОШИНОВСЬКА І.А.
(Національний університет "Львівська політехніка")

ОСОБЛИВОСТІ АВТОРИЗАЦІЇ ВУЗЬКОСПЕЦІАЛІЗОВАНИХ НАУКОВИХ ПРАЦЬ

1. Вступ

Інформація, представлена у текстовій формі, відображає далеко не лише зміст явно викладеної автором ідеї. Кожна авторська праця характеризує самого автора або ж авторський колектив, даючи змогу відтворити найбільш вагомі фактори, що мали вплив на процес написання роботи. Аналіз особливостей використання автором

синтаксичних конструкцій та граматичних форм, оцінка об'єму лексичного словника автора, середньої кількості слів у реченні та кількості речень в абзаці є базовими процедурами процесів ідентифікації та діагностики автора тексту [Марусенко, 1990], [Виноградов, 1961], [Батов, 1977], [Статистичні параметри стилів., 1967]. Можливість ефективного застосування методу аналізу головних компонент для авторизації тексту показано в роботах [Burrows, 1992], [Burrows, 2003]. Автори [Baayen, 2002], [Juola, 2003] продемонстрували успішну класифікацію робіт за тематикою, автором, його віковою категорією та освітою, використовуючи метод лінійного дискримінантного аналізу з врахуванням ентропії частотних характеристик слів в межах досліджуваного набору текстів. Використання теорії графів в процесі автоматичної обробки текстів з метою встановлення зв'язків між словами для стилістичної діагностики тексту обговорено у працях [Севбо, 1981], [Скороходько, 1983], [Скороходько, 2006]. Дослідження особливостей вираження ділового стилю проведено в роботі [Пещак, 1980].

Процедура авторизації ускладнюється у випадку текстів з галузі природничих та точних наук, де авторський стиль не так вже й легко ідентифікувати. Автор наукової статті досить обмежений в стилі та засобах вираження своїх думок, він змушений використовувати вже усталену термінологію, висловлювати свої ідеї та результати точно, стисnуто, однозначно. Авторизація наукових праць є особливо актуальною для розпізнавання плагіату, що вкрай необхідно для захисту авторських прав за умов стрімкого науково-технічного прогресу. Нажаль, в практиці доволі часто доводиться з'ясовувати хто, коли, що і в кого «запозичив».

Зазвичай, наукова стаття, особливо експериментального характеру, відображає результати колективної роботи куратора, безпосередніх виконавців експерименту, консультантів. Авторизація таких робіт є більш однозначною за умови прояву лексикологічних особливостей одного із співавторів статті. Формування структури статті, та внесення поправок до вже написаного відбувається згідно результатів колективного обговорення. Ключовим моментом, що визначає можливість подальшої авторизації є остаточне редагування статті, яке проводиться, здебільшого, одним із співавторів з врахуванням зауважень та уточнень колективу. Слід мати на увазі можливість зміни стилю викладення матеріалу статті відповідно до коментарів від рецензентів, а також внесення редакторських правок згідно з вимогами журналу. Всі ці чинники ускладнюють процес авторизації наукової статті.

В даній роботі проводиться пошук оптимальних параметрів процесу ідентифікації авторського стилю методом аналізу головних компонент N-грам моделі статей вузькоспеціалізованого спрямування.

2. Об'єкт та методика дослідження

В роботі проаналізовано 16 статей, присвячених тематиці люмінесцентної спектроскопії. Усі проаналізовані статті є результатом колективних експериментальних досліджень із спільним співавтором Стриганюком Г.Б., опублікованих або поданих до друку англійською мовою у реферованих журналах (див. Табл. 1). Статті поділено на дві групи: група А представлена статтями, в яких даний співавтор приймав лише участь в обговоренні, але не проводив кінцевої правки тексту статті; група В поєднує статті, остаточне редагування яких проводив досліджуваний автор. Аналіз робіт погоджено з Стриганюком Г.Б., який попередньо провів класифікацію статей на дві групи (А і В). Статті позначено латинськими літерами відповідно до груп та пронумеровано для наглядної візуалізації їх розподілу.

Провівши статистичну обробку вибраного масиву з 16 текстів, створено словник, що налічує 5000 словоформ та сформовано частотну матрицю **A**(5000×16), елемент a_{ij} якої відображає абсолютну частоту появи *i*-ого слова в *j*-ому тексті. Для аналізу

головних компонент формувалась транспонована матриця $\mathbf{A}^T$ розмірністю 16×1897 (текстів×слів), в якій не брались до розгляду частоти стоп-слів та словоформ, що зустрілись лише в одному тексті.

В роботі досліджуються функціональні можливості поєднання методу аналізу головних компонент [Pearson, 1901], [Jolliffe, 2002] з N-грам моделями текстів [Cavnar, 1994] у атрибуції авторського стилю наукових праць. Метод аналізу головних компонент полягає у виборі системи координат для оптимального представлення багатомірних даних. Осями такої координатної системи вибираються власні вектори $\mathbf{p}_i$ коваріаційної матриці:

$$\mathrm{cov}(\mathbf{A}) \cdot \mathbf{p}_i = \lambda_i \cdot \mathbf{p}_i , \qquad (1)$$

де λ_i – власні значення коваріаційної матриці, яка формується наступним чином:

$$\mathrm{cov}(\mathbf{A}) = (\mathbf{A}^T \cdot \mathbf{A})/(m\text{-}1), \qquad (2)$$

m – кількість стрічок транспонованої матриці $\mathbf{A}^T$ (кількість досліджуваних текстів). В рамках N-грам моделі використовується припущення про залежність ймовірності появи n-ого слова в реченні від комбінації попередніх n-1 слів. При цьому, ймовірність формування певного речення оцінюється як добуток ймовірностей усіх N-грам, які входять до його складу. Найбільш вдалою для англомовних текстів вважається 3-грам модель [Lafferty, 1992].

Таблиця 1. Список проаналізованих статей з відповідними їм позначеннями

Група А

A1 P.A.Rodnyi, G.B.Stryganyuk, C.W.E. van Eijk et al., Variation of 5*d*-level position and emission properties of BaF_2:Pr crystals, *Phys. Rev. B* **72**, *19* (2005) 195112

A2 Ya.Chornodolskyy, G.Stryganyuk, S.Syrotyuk et al., Features of the core-valence luminescence and electron energy band structure of $A_{1-x}Cs_xCaCl_3$ (A=K, Rb) crystal, *J. Phys.: Condens. Matter* **19** (2007) 476211

A3 P.A.Rodny, A.Kh.Khadro, A.S. Voloshinovski, and G.B.Stryganyuk, Europium Luminescence in Fluorite upon High-Energy Excitation, *Optics and Spectroscopy*, **103**, *4* (2007) 568

A4 V.N.Makhov, S.Kh.Batygov, L.N.Dmitruk, M.Kirm, G.Stryganyuk, G.Zimmerer, VUV 5d-4f luminescence of Gd^{3+} and Lu^{3+} ions doped into CaF_2, *Phys. Stat. Sol. (c)* **4**, *3* (2007) 881

A5 M.Kirm, G.Stryganyuk, S.Vielhauer et al., Vacuum-ultraviolet 5d-4f luminescence of Gd^{3+} and Lu^{3+} ions in fluoride matrices, *Phys. Rev. B* **75** (2007) 75111.

A6 N.V.Guerasimova, I.A.Kamenskikh, V.V.Mikhailin, C.Pedrini, A.Petrosyan, G.Stryganyuk and G.Zimmerer, Intrinsic luminescence and luminescence of inadverten impurities in LuAP and LuYAP crystals, *Rus. Phys. J.* **49** (2006) 44

A7 E.Savchenko, A.Ogurtsov, I.Khyzhniy, G.Stryganyuk, G.Zimmerer, Creation of permanent lattice defects via exciton self-trapping into molecular states in Xe matrix, *Phys. Chem. Chem. Phys.* **7** (2005) 785

Група В

B1 A.Voloshinovskii, G.Stryganyuk, G.Zimmerer et al., Luminescent characteristics of pure and Ce doped K_2LaCl_5 phase in KCl host., *Phys. Stat. Sol. (a)* **202**, *9* (2005) R101

B2 G.B.Stryganyuk, P.V.Savchyn, Z.A.Khapko et al., Luminescence of Pr^{3+} doped K_2LaCl_5 microcrystals encapsulated in KCl host, *Opt. Mater.* (2008), doi:10.1016/j.optmat.2008.07.001

B3 A.S.Voloshinovskii, O.T.Antonyak, G.B.Stryganyuk et al., Spectral-kinetic study of $LaCl_3$:Ce crystals, *Ukr. J. Phys. Opt.* **3**, *3* (2002) 194

B4 G.Stryganyuk, D.Trots, A.Voloshinovskii et al., Luminescence of Ce^{3+} doped $LaPO_4$ nanophosphors upon Ce^{3+} *4f-5d* and band-to-band excitation, *J. of Lumin.* **128** (2008) 355

B5 G.Stryganyuk, G.Zimmerer, N.Shiran et al., Spectral-kinetic characteristics of Pr^{3+} luminescence in $LiLuF_4$ host upon excitation in the UV–VUV range, *J. Lumin.* **128** (2008) 1937

B6 G.Stryganyuk, S.Zazubovich, A.Voloshinovskii et al., Charge transfer luminescence of Yb^{3+} ions in $LiY_{1-x}Yb_xP_4O_{12}$ phosphates, *J. Phys.: Condens. Matter* **19** (2007) 036202

B7 V.V.Vistovskyy, P.V.Savchyn, G.B.Stryganyuk et al., Luminescence of Ce doped $LaCl_3$ microcrystals incorporated into a single-crystalline NaCl host, *J. Phys.: Condens. Matter* **20** (2008) 325218

B8 A. Voloshinovskii, I. Pashuk, Ya. Chornodol'skyi, G. Stryganyuk, P. Rodnyi. Peculiarities of core-valence luminescence of CsBr and $Rb_{1-x}Cs_xBr$ crystals, *Phys. Stat. Sol. (b)* **241**, *11* (2004) 2613

B9 G.Stryganyuk D.Trots, I.Berezovskaya et al., Luminescence of YbP_3O_9 upon excitation in the UV–VUV range, *J. Phys.: Condens. Matter* **19** (2007) 346236

Для покращення ідентифікації авторського стилю проведено аналіз головних компонент моделі N-грам (N=2, ..., 6). В таблиці 2 представлена розмірність частотних матриць **A**, сформованих для N-грам моделей розглянутих текстів. До розгляду брались N-грами (N=1, ..., 6), що зустрілись у 2-ох та більше текстах.

Таблиця 2. Кількість N-грам, виявлених та прийнятих до розгляду при аналізі головних компонент масиву з 16 текстів.

N-грам	Виявлено	Прийнято до розгляду
1-грам	5000	1897
2-грам	20565	4532
3-грам	30694	3300
4-грам	33287	1733
5-грам	32007	878
6-грам	29301	492

3. Результати та обговорення

Для аналізу головних компонент було взято транспоновану матрицю A^T розмірністю 16×1897. Модель головних компонент, побудована на основі аналізу частотних характеристик словоформ, що зустрілися у двох і більше текстах, описує 100% варіації змінних (словоформ) за допомогою 16 головних компонент. Для аналізу розподілу текстів та словоформ відповідно до авторського стилю до розгляду взято 10 перших головних компонент, що описують 92,13% варіації змінних. Вклад перших чотирьох компонент у модель складає 73,18%. Поєднання другої, третьої та четвертої головних компонент, представлених на Рис.1 як PC 2, PC 3, PC 4 (Principal Components) виявилось найбільш інформативним базисом для побудови просторової моделі досліджуваного набору текстів. Перша головна компонента PC 1 не була прийнята до розгляду, оскільки словоформи, які дають найбільший вклад у її формування, є характерними для опису властивостей, що притаманні усім об'єктам, окресленим у досліджуваних текстах.

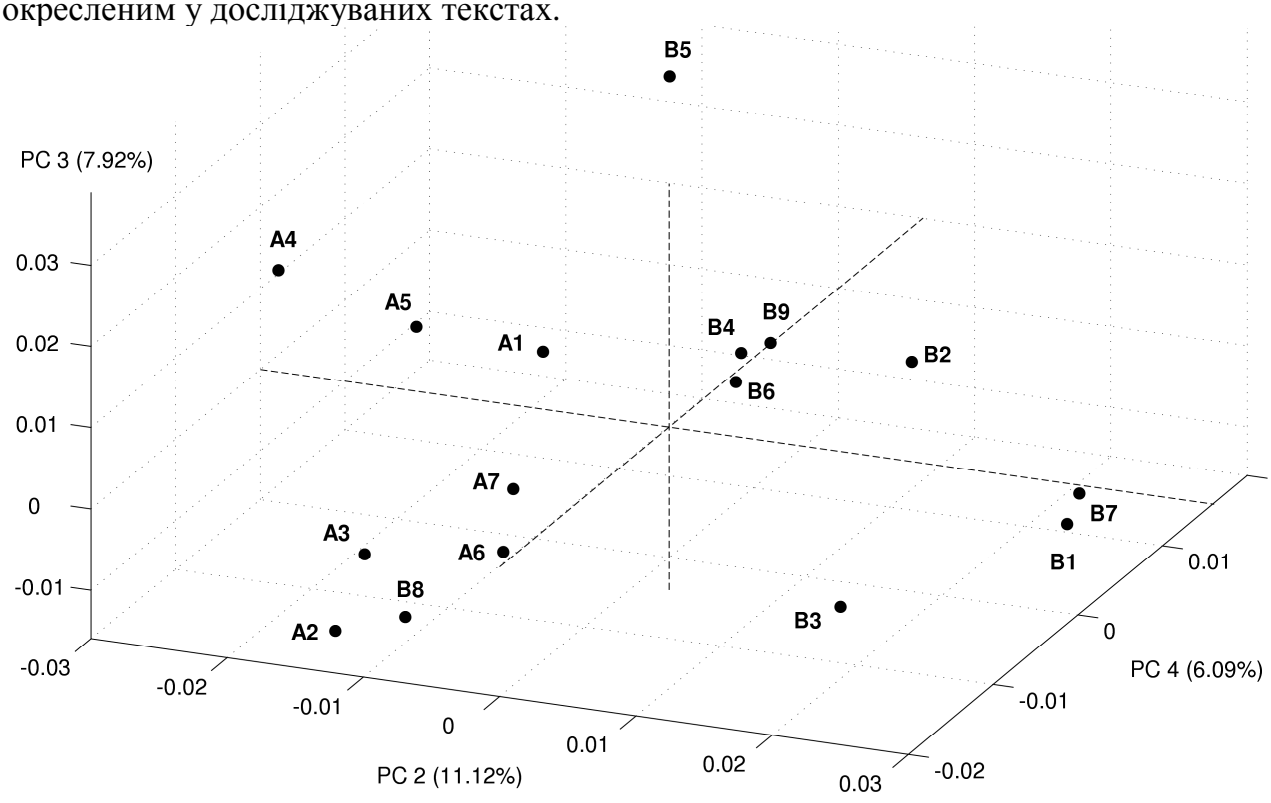


Рис. 1. Представлення текстів у просторі головних компонент PC 2, PC 3, PC 4 для моделі 1-грам.

Розглянуті статті розділились на дві групи в просторі головних компонент PC 2, PC 3, PC 4. Друга головна компонента PC 2 відповідає характеристикам, які в найбільшій мірі розділяють статті за авторським стилем. Роботи, редаговані досліджуваним автором (група B), вдалось чітко виділити з поміж усіх робіт за його співавторством. Виняток складає стаття B8, яка розмістилась в просторі головних компонент поблизу статті A2. Роботи A2 та B8 присвячені одній тематиці і їх близьке розташування вказує на суттєвий вклад тематичної термінології у головні компоненти PC 2, PC 3, PC 4.

Рисунок 2 підтверджує значний вклад тематичної термінології у формування головних компонент моделі. Таким чином, розподіл статей, представлений на Рис. 1, в значній мірі обумовлений спорідненістю об'єктів досліджень. Наприклад, близьке розташування статей B1, B7, B3 відносно напрямку PC 2 на Рис. 1 визначається ключовими словами $LaCl_3$, Ce, K_2LaCl_5 (Рис. 2), що відображають об'єкти досліджень даних робіт. Щоб уникнути групування робіт за об'єктом досліджень, були вилучені назви досліджуваних об'єктів (251 словоформ) з частотної матриці **A**. Розмірність матриці $\mathbf{A}^T$ була таким чином зменшена від 16×1897 до 16×1646. Перерахована після цього модель 10-и головних компонент описує 91,94% частотної варіації словоформ, причому, 75,52% варіації припадає на перші чотири компоненти.

Після вилучення з розгляду назв сполук та методів галузі, розрахована для 1-грам модель головних компонент забезпечує не лише групування в просторі PC 2, PC 3, PC 4 (див. Рис. 3) статей, редагованих Г. Стриганюком, але вказує на лексичну близькість робіт, авторська викладка яких належить професору Санкт-Петербурзького політехнічного університету П.О. Родному (роботи A1, A3) та науковцям естонської дослідницької групи М. Кірма з університету Тарту (роботи A4, A5).

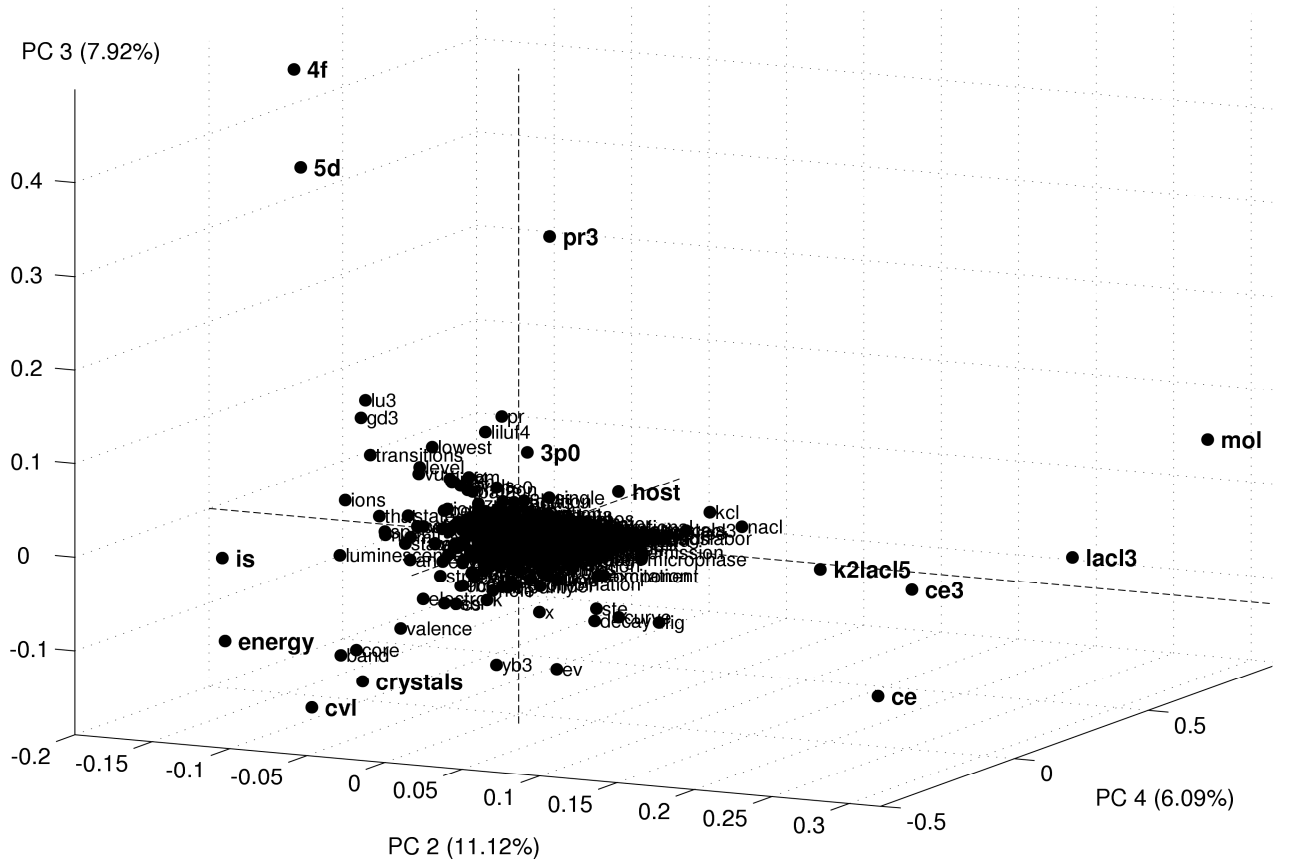


Рис. 2. Вклад слів у формування головних компонент PC 2-4 моделі 1-грам.

Стаття B8, все-таки, залишилась поміж статей групи A в просторі головних компонент, незважаючи на виключення з розгляду тематичної термінології. B8 є однією з перших робіт у співавторстві досліджуваного автора, тому відокремлення роботи B8 від групи B можна пояснити становленням його авторського стилю. Фактор становлення та зміни наукового стилю є особливо важливим у випадку написання робіт іноземними мовами.

Покращення ідентифікації авторського стилю в роботі реалізовано шляхом залучення аналізу характерних автору лексичних конструкцій, для чого моделі головних компонент були розраховані для N-грам (N=2, ..., 6). Найбільш чіткої категоризації статей за стилем автора було досягнуто у моделі головних компонент для 4-грам (Рис. 4). Слід зауважити, що кількість виявлених елементів N-грам моделі сягає максимуму саме для 4-грам (див. Табл. 2).

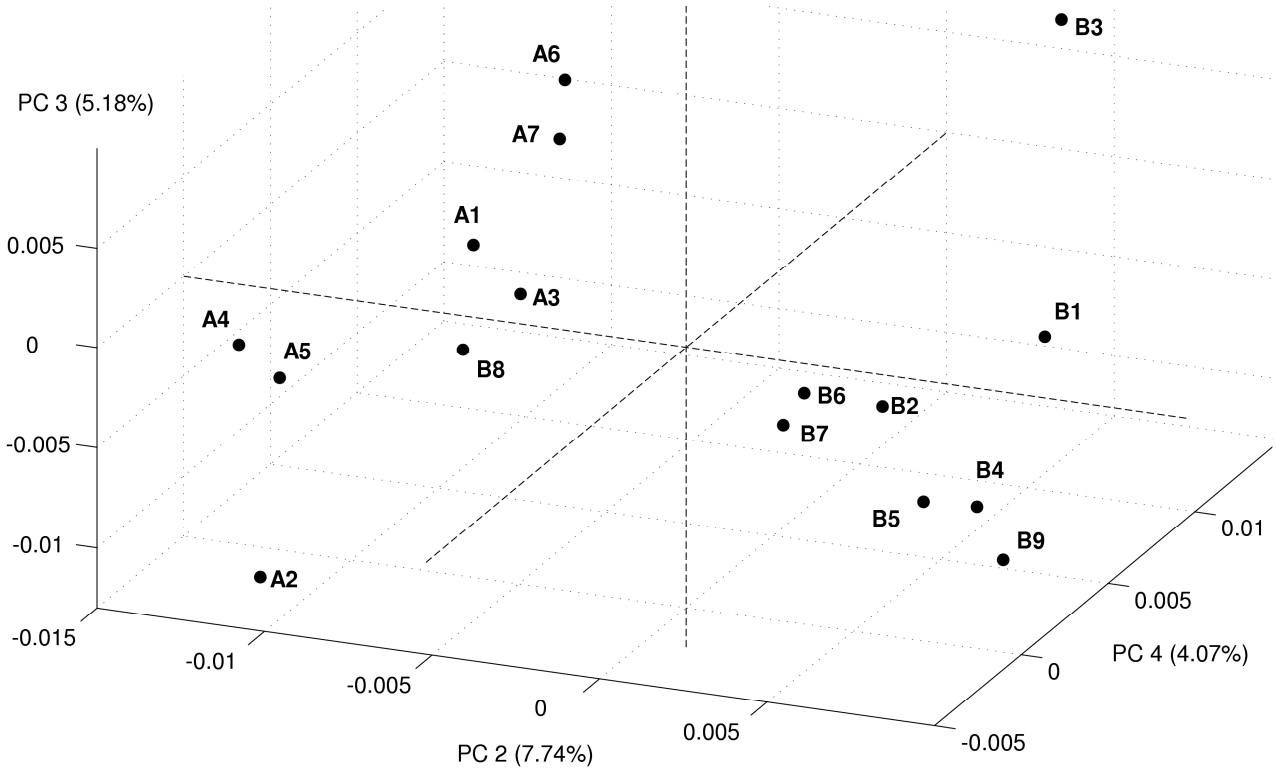


Рис. 3. Представлення текстів у просторі головних компонент PC 2-4 для моделі 1-грам, розрахованої без врахування тематичної термінології.

Наглядне розділення між роботами груп A та B виявлено за ознаками першої компоненти, опис яких вкладається у 23,19% 4-грам моделі. Аналіз вкладу 4-грам у формування першої компоненти дає змогу виділити лексичні особливості стилю робіт групи B. Прояви стилю досліджуваного автора виявленні у використанні трьох типів 4-грам, характерних для: 1) опису спостережуваних об'єктів; 2) вираження припущень; 3) представлення та порівняння результатів (Табл. 3). Слід зауважити властиве автору вживання припущень, про що свідчить поширеність модальних слів *may* та *can* у виявлених характерних автору 4-грамах.

Таблиця 3. 4-грами, притаманні авторському стилю досліджуваного автора.

опис об'єктів	вираження припущень	представлення результатів
have been found to	one can expect the	one may conclude the
is revealed for the	may correspond to the	is estimated to be
is more pronounced for	may be caused by	estimated to be about
appears due to the	may be considered as	a good agreement with
upon the excitation in	may be attributed to	

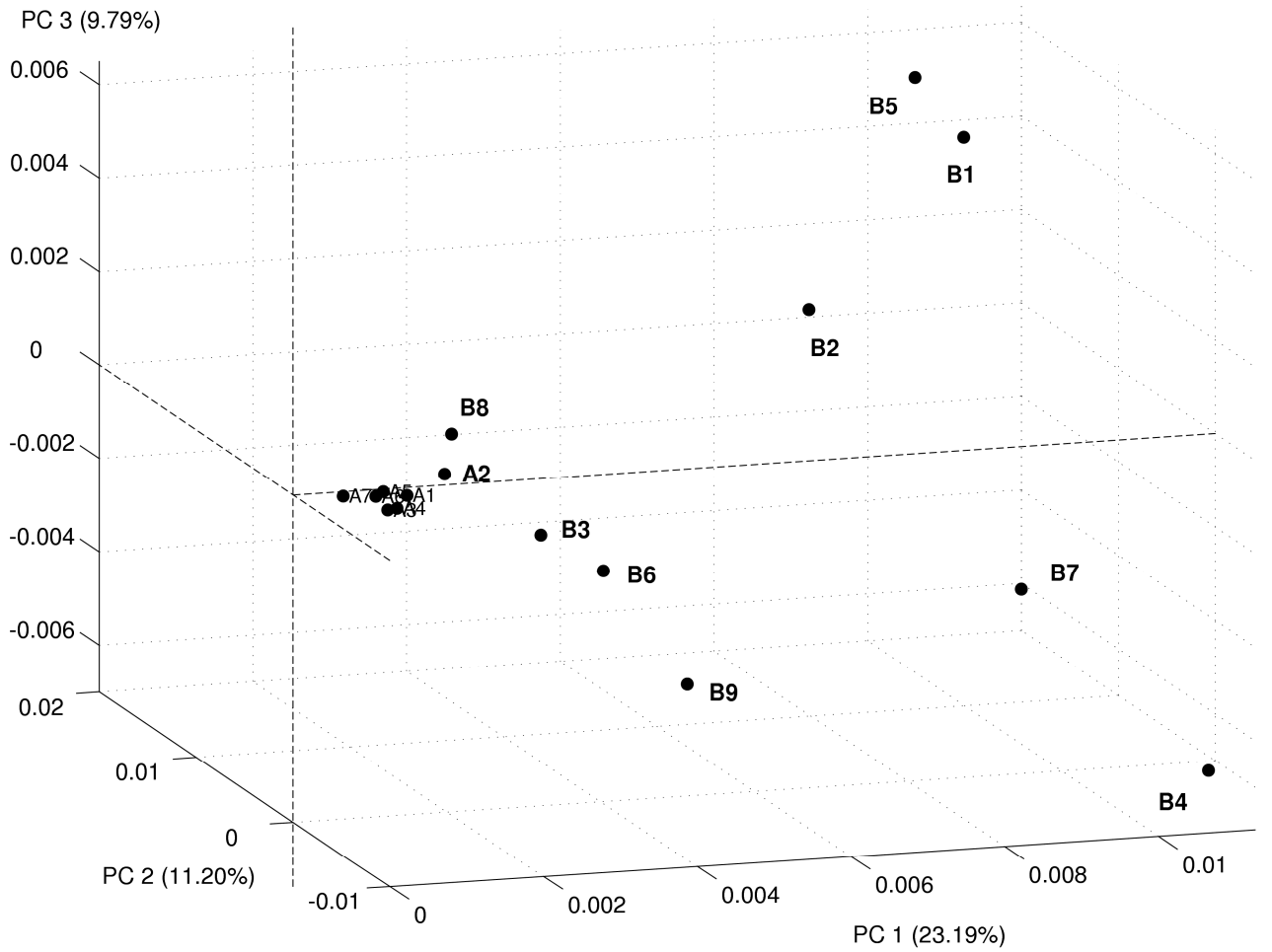


Рис. 4. Представлення текстів у просторі головних компонент PC 1-3 для моделі 4-грам.

5. Висновки

Застосування методу аналізу головних компонент до частотної матриці словоформ текстового масиву дозволяє провести авторську атрибуцію вузькоспеціалізованих робіт без апріорних даних щодо стилю автора. Для ефективної ідентифікації авторського стилю доцільно розглядати N-грам моделі текстів. Найкраще розділення авторського стилю слід очікувати при аналізі N-грам, розмір яких забезпечує їх найбільшу кількість у тексті.

Аналіз головних компонент N-грам моделі дає змогу виявити характерні лексичні набори, властиві певному авторському стилю. Найбільш характерні лексичні набори автора, виявлені методом аналізу головних компонент N-грам моделі, можна використовувати для пошуку робіт шляхом ідентифікації авторського стилю.

Література

Батов В.И., Сорокин Ю.А. "Опыт построения методики для установления авторства текстов" – Изв. АН СССР, сер. Лит. и языка. – Т. 36. – №4. – 1977. *Виноградов* В.В., Проблема авторства и теория стилей. – М.: Государственное издательство Художесвенной литературы, 1961. – 615 с. *Марусенко* М.А., Атрибуция анонимных и псевдонимных литературных призведений методами распознавания образов. – Л.: Издательство Ленинградского университета, 1990. – 168 с. *Пещак* М.М., Стиль деловых документов XIV века, Автореф. дис. д-ра філол. наук. – К., 1980. – 51 с. *Севбо* И.П., Графическое представление синтаксических структур и стилистическая диагностика. – К.: Наукова Думка, 1981, 372 с. – *Скороходько* Е.Ф., Семантические сети и автоматическая обработка текста. – К.: Наукова Думка, 1983. 220 с. – *Скороходько* Е.Ф., Термін у науковому тексті (до створення

терміноцентричної теорії наукового дискурсу). – К.: Логос, 2006. – 99 с. – *Статистичні* параметри стилів. Відп. ред. Перебийніс В.С., – К.: Наукова Думка, 1967. – 260 с. *Baayen* H, H. van Haltern, A. Neijt, F. Tweedie, "An experiment in authorship attribution," in Proc. of JADT 2002, St.Malo, pp.29-37. *Burrows* J., "Not unless you ask nicely: The interpretative nexus between analysis and information," Literary and Linguistic Computing, vol. 7, 1992, pp. 91-109. *Burrows* J., "Questions of Authorships: Attribution and Beyond," Computers and the Humanities, vol. 37, No.1, 2003, pp.5-32. *Cavnar* W.B., Trenkle J.M., "N-Gram-Based Text Categorization," in Proc. of 3rd Annual Symposium on Document Analysis and Information Retrieval, Las Vegas, 1994, pp. 161-175. *Jolliffe* I.T., Principal Component Analysis (Springer Series in Statistics) 2nd ed., NY: Springer, 2002, 487 p. *Juola* P., "The time Course of language change," Computers an the Humanities, vol. 37, No.1, 2003, pp.77-96. *Lafferty* J.D., Sleator D., Temperley D., "Grammatical trigrams: A probabilistic model of link grammar, " in Proc. of AAAI Fall Symp. Probabilistic Approaches to Natural Language, Cambridge, MA, Oct. 1992, pp. 74-81. *Pearson* K., "On Lines and Planes of Closest Fit to Systems of Points in Space". Philosophical Magazine, vol. 2, 1901, pp.559-572.

ВАРАКІНА В.В.
(*Національний тех. ун-т ХПІ*)

ДИСКУРС ЯК СОЦІАЛЬНО ОБУМОВЛЕНА ФОРМА МОВЛЕННЄВОЇ ВЗАЄМОДІЇ

Одним із центральних понять сучасної лінгвістики є дискурс, що представляє собою складне та багатогранне явище, яке досліджується з позицій різних напрямків лінгвістики: прагмалінгвістики, соціолінгвістики, психолінгвістики, когнітивної лінгвістики, етнолінгвістики, лінгвістики тексту і т. і. з залученням психології, соціології, філософії та інших наук.

Широкий діапазон дискурсивних досліджень у сучасній лінгвістиці демонструє багатозначність терміну дискурс та його міждисциплінарний характер. Запропонований 3. Харрісом у 1952 році термін „дискурс" у значенні структурної одиниці мови, яку можна вивчати так само як і речення, протягом півсторіччя зазнав значних доповнень, переосмислень і модифікацій.

Тому **мета** даної роботи вивчення напрямів дискурсивних досліджень у сучасній лінгвістиці. На сьогоднішній день чіткого та загальновизнаного визначення дискурсу, що охоплює усі випадки його вживання, не існує - різноманітні тлумачення цього терміну успішно задовольняють різноманітні понятійні потреби, привносячи дещо нове у традиційне уявлення про мовлення, текст, діалог, стиль та мову.

Завдання даної роботи звести до купи різноманітні існуючі підходи до вивчення дискурсу. **Предмет** дослідження об'ємний функціональний потенціал дискурсу. **Об'єкт** дослідження дискурс як соціально обумовлена форма мовленнєвої взаємодії.

У цілому існуючі підходи до вивчення дискурсу можна звести до таких:

1.Дискурс - "зв'язний текст у сукупності з екстралінгвістичними, соціокультурними, прагматичними, психологічними та іншими факторами; це - текст, узятий в аспекті подій; мовлення, що розглядається як цілеспрямоване соціальне явище, дія, як компонент, що бере участь у взаємодії людей та механізмах їх свідомості. Дискурс - це мовлення, заглиблене в життя " [Арутюнова 1990, С.136-137].

2.Дискурс визначається через текст, або текст - через дискурс[Зільберт 1991, с.56].