

УДК 81'32:81'367:81'373.7

**СИТАР Ганна**

## **ОБЧИСЛЕННЯ ПОКАЗНИКА АСОЦІАЦІЇ *MI LOG FREQ* ЯК СТАТИСТИЧНИЙ МЕТОД ДОСЛІДЖЕННЯ СИНТАКСИЧНИХ ФРАЗЕОЛОГІЗМІВ**

Статтю присвячено статистичному аналізу синтаксичних фразеологізмів на матеріалі української мови. Здійснено обчислення показника асоціації  $MI \log Freq$ , який є модифікованою формулою коефіцієнта  $MI$  (mutual information) та дає змогу виявити ступінь не випадковості (статистично доведеної зв'язаності) поєднання словоформ, що входять до складу незмінного компонента моделі речення.

Подано результати виконаних обчислень для низки моделей синтаксичних фразеологізмів української мови за даними Українського національного лінгвістичного корпусу.

**Ключові слова:** конструкція, корпус текстів, синтаксичний фразеологізм, статистичний аналіз, показник асоціації, фразеологізоване речення.

Синтаксичні фразеологізми, або фразеологізовані речення, – специфічний тип речення, у якому фіксовано розташовані постійний (незмінний) і змінний компоненти пов'язані ідіоматично, граматичні зв'язки і прямі лексичні значення слів послаблені або втрачені на сучасному етапі розвитку мови. Синтаксичні фразеологізми є дієвим засобом вираження ставлення мовця до висловлюваного, властивим для розмовного мовлення, художніх і публіцистичних текстів [1; 2; 4 та ін.] (докладно про ознаки та статус синтаксичних фразеологізмів у системі мовних одиниць див. у праці [5]).

Актуальність дослідження зумовлена потребою здійснення статистичного аналізу синтаксичних фразеологізмів на матеріалі корпусу текстів як значного і реперезентативного мовного матеріалу, що разом із вибором адекватних статистичних показників, застосовуваних до відповідних типів мовних одиниць, забезпечує вірогідність отриманих даних.

Українська лінгвостатистика має значні здобутки у статистичному аналізі лексем, груп слів, текстів певних стилів – наукового, публіцистичного, художнього та ін. (див. праці

В. Перебийніс, Н. Дарчук, В. Левицького, С. Бук та ін. дослідників), проте майже не звертається до статистичного аналізу різноманітних поєднань слів (колокацій, конструкцій та ін.).

Виконання теоретичної частини нашого дослідження дало змогу сформулювати таку робочу гіпотезу: синтаксичні фразеологізми, як і будь-які інші стійкі одиниці, мають високий ступінь не випадковості поєднання слів, що входять до складу незмінної частини речення. Спробуємо довести правильність цього твердження за допомогою обчислення одного з показників асоціації – MI Log Freq.

Предметом пропонованого статистичного аналізу стали синтаксичні фразеологізми, незмінний компонент яких складається з поєднання кількох лексем – службових і повнозначних слів, яким властиве семантичне спустошення або семантичний зсув (*що за, от тобі/вам і/й, не до, теж мені, яке там і под.*): **Що за пісня! От вам і подяка! Теж мені друг! Яке там встигла! Реформи і в Африці реформи.**

Мета дослідження – виявити ступінь не випадковості словоформ, що входять до незмінного компонента синтаксичних фразеологізмів за допомогою обчислення показника асоціації MI log Freq. Поставлена мета передбачає розв'язання таких завдань: 1) розглянути сутність показника асоціації MI log Freq; 2) здійснити обчислення за формулою MI log Freq для низки моделей синтаксичних фразеологізмів за даними Українського національного лінгвістичного корпусу; 3) зіставити отримані результати з відповідними даними, одержаними за вихідною формулою MI.

За Кембріджським словником статистики Брайана Еверітта (Brian S. Everitt), показники асоціації (англ. association measures, measures of association) – «числові індекси, що обчислюють силу статистичної залежності двох або більше квалітативних змінних» [10, р. 241] (переклад мій – Г.С.).

У лінгвістичній статистиці останніх десятиліть застосовують кілька десятків показників асоціації (див. [11; 14; 16] та ін.). Для вивчення колокацій / конструкцій найчастіше використовують коефіцієнти MI (або MI-score), t-score, log-likelihood, Dice, gmean та ін. На попередньому етапі статистичного дослідження ми

обґрунтували доцільність застосування показника асоціації МІ до аналізу синтаксичних фразеологізмів [6].

Поняття МІ (англ. mutual information – взаємна, спільна, повна інформація) запропоноване в теорії інформації Робертом Маріо Фано (Robert Mario Fano) [13]. У лінгвістиці це поняття застосували Кеннет Ворд Чарч (Kenneth Ward Church) та Патрік Хенкс (Patrick Hanks) для вивчення випадковості / не випадковості поєднання двох лексем, тобто біграм [9].

Оскільки значна частина синтаксичних фразеологізмів є багатокомпонентними одиницями, постає потреба в урахуванні більшої кількості компонентів, ніж два, тому ми здійснюємо обчислення МІ за формулою (1), яка дає змогу врахувати будь-яку кількість слів (і ≥ 2). Формулу (1) подаємо за [8, с. 586]:

$$MI = \log_2 \frac{f(c_1, c_2, \dots, c_i) \times N^{(i-1)}}{f(c_1) \times f(c_2) \times \dots \times f(c_i)}, \quad (1)$$

де МІ – коефіцієнт mutual information; і – це кількість компонентів конструкції;  $c_1$  – перша лексична одиниця;  $c_2$  – друга лексична одиниця;  $c_i$  – і-а лексична одиниця;  $f(c_1, c_2, \dots, c_i)$  – абсолютна частота вживання конструкції  $c_1, c_2, \dots, c_i$  в корпусі (з урахуванням порядку одиниць усередині конструкції);  $f(c_1)$  – абсолютна частота  $c_1$  у корпусі;  $f(c_2)$  – абсолютна частота  $c_2$  в корпусі;  $f(c_i)$  – абсолютна частота  $c_i$  в корпусі;  $N$  – загальна кількість слів у корпусі;  $\log_2$  – логарифм числа за основою 2.

На нашу думку, статистичний критерій МІ відповідає таким вимогам: дає змогу визначити коефіцієнт не випадковості поєднання двох і більше слів у корпусі текстів, враховує частоту конструкції, частоту її компонентів, розмір корпусу та має формулу в узагальненому вигляді для конструкцій з будь-якою кількістю компонентів [6].

Як свідчить досвід обчислення МІ для різних типів колокацій / конструкцій на матеріалі англійської, німецької та російської мов, показник МІ виявляється не виправдано великим для поєднань слів з дуже низькою частотою (наприклад, частотою 1) [3; 7; 8; 11; 16; 17]. Для того, щоб уникнути підвищення значущості таких конструкцій, учені вводять кілька модифікованих формул

коефіцієнта  $MI - MI^2, MI^3, MI \text{ Log Freq}$ , *salience* та ін. Усі вони в той чи інший спосіб підвищують значущість частоти конструкції, а не окремих її компонентів.

На позначення модифікованих, похідних формул Штефан Еверт (Stefan Evert) уживає термін «евристичні формули» (грецьк. *ευρίσκειν* (*heuristiko*) – знаходжу, відшукую, відкриваю) і виділяє їх як окрему групу показників асоціації, застосування яких є виправданим з урахуванням специфіки предмета і мети статистичного дослідження [11, р. 77, 89-91].

Правомірність обчислень за «альтернативними», тобто модифікованими, формулами Павел Пеціна (Pavel Pecina) та Вайолета Серетан (Violeta Seretan) обґрунтовують прагненням одержати оптимальні кількісні результати щодо одиниць певного типу у специфічних умовах [14; 16]. Вибір показника асоціації зумовлений передусім такими чинниками, як мета дослідження, тип мовної одиниці, розмір корпусу й тип тексту [12, р. 452; 16, р. 42 і далі].

Показник асоціації  $MI \text{ log Freq}$  отримують шляхом множення коефіцієнта  $MI$  на логарифм частоти конструкції. Відповідно в назві цього коефіцієнта відбито сутність запропонованої модифікації:  $MI \text{ log Freq}$  – *mutual information logarithm frequency*. Формулу (2), використану в одній із версій корпусного програмного забезпечення Sketch Engine (<https://www.sketchengine.co.uk>), наводимо за працею Павела Рихли (Pavel Rychlэ) [15, р. 8]:

(2)

$$MI \text{ log Freq} = MI - score \times \log f_{xy},$$

де *MI-score* – показник взаємної інформації;  $f_{xy}$  – абсолютна частота вживання біграми  $xy$  в корпусі; *log* – логарифм числа за будь-якою основою.

Під час створення комп'ютерної програми, призначеної для автоматичного виділення двокомпонентних колокацій (біграм) в російськомовному тексті Віра Залеська використовує комбінацію з дев'яти показників асоціації [3]. Одним із цих показників є коефіцієнт, позначений терміном «*salience*» (англ. *salience* – виступ, опуклість), або  $MI \text{ log Freq}$ , та кваліфікований як «нормалізований

варіант показника  $MI$ », що «збільшує вагу частоти спільного вживання ключового слова та його колоката порівняно з  $MI$ . Отже, ефективність *salience* [щодо автоматичного визначення колокації – *Г.С.*] ... має бути вищою, ніж у  $MI$ » [3, с. 285] (переклад мій – *Г.С.*). Дослідниця наводить формулу (3):

$$salience = MI \times \ln (f(n, c) + 1), \quad (3)$$

де  $MI$  – показник взаємної інформації;  $\ln$  – натуральний логарифм;  $f(n, c)$  – абсолютна частота вживання біграми  $nc$  у корпусі.

Ми здійснюємо обчислення за формулою (3), оскільки вона має такі переваги: а) комп'ютерні програми використовують натуральні логарифми (логарифми за основою  $e$  – ірраціональною константою, що дорівнює приблизно 2,718281828), відповідно вибір саме цієї формули вмотивований можливим подальшим її вмонтуванням до корпусу з метою автоматичного здійснення підрахунків; б) до частоти конструкції додано 1 з метою уникнення програмної помилки, якщо  $f(n, c) = 0$  (оскільки  $\log 0$  не існує). Цю формулу позначаємо терміном « $MI \log \text{Freq}$ » як цілком зрозумілим і прозорим.

Статистичний аналіз синтаксичних фразеологізмів ми здійснювали за даними Українського національного лінгвістичного корпусу (далі УНЛК), створеного колективом Українського мовно-інформаційного фонду НАН України та розміщеного за адресою [http://unlc.icybcluster.org.ua/virt\\_unlc/](http://unlc.icybcluster.org.ua/virt_unlc/)<sup>1</sup>. Для встановлення абсолютної частоти конструкції та абсолютної частоти окремих словоформ, що входять до складу конструкції, в пошуковій формі корпусу було передбачено пошук словоформи, а не слова з урахуванням його парадигми, та задано визначений порядок словоформ. Загальна кількість слововживань в УНЛК у жовтні 2015 року становила 180 мільйонів слововживань.

---

<sup>1</sup> Дякуємо Директору Українського мовно-інформаційного фонду НАН України академіку НАН України Володимиру Широкову за наданий доступ до корпусу.

Коефіцієнти MI та MI log Freq обраховували з точністю до двох знаків після коми. Фрагмент отриманих результатів для 10 різнотипних синтаксичних фразеологізмів наведено в таблиці 1. Через скісну ризку подано можливі варіанти в межах однієї моделі, дужками позначено факультативність компонента моделі. Показники асоціації обраховано окремо для кожного варіанта моделі.

Як видно з таблиці 1, показник асоціації MI log Freq для всіх обстежених моделей синтаксичних фразеологізмів є високим, перебуває в межах від 44,99 (*оце ... так*) до 296,11 (*що ж це за*). Проте, на відміну від коефіцієнта MI, у випадку обчислення показника MI log Freq не зафіксовано статистично вірогідного зв'язку між кількістю компонентів конструкції та його величиною (пор.: для синтаксичних фразеологізмів із двочленним незмінним компонентом MI перебуває в межах від 10,77 (*оце ... так*) до 14,31 (*що за*), для трикомпонентних моделей – від 22,09 (*яке ж там*) до 28,4 (*де вже там*), для чотирикомпонентних моделей – від 38,32 (*вони і в Африці*) до 41,27 (*що ж це за*)). Випадки, коли значення MI log Freq для трикомпонентних конструкцій є нижчим, ніж у двокомпонентних (зокрема, *яке ж там* 45,03 і *яке там* 61,67), або значення чотирикомпонентних є нижчим за трикомпонентні (наприклад, *він і в Африці* 68,82, *вона і в Африці* 68,74, *вони і в Африці* 52,08 та *і в Африці* 88,56) пояснюються низькою частотою конструкцій з більшою кількістю складників (у наведених прикладах відповідно 7 і 208; 5, 5, 3 і 30).

Отже, отримані кількісні дані за формулами MI та MI log Freq підтверджують висунуту гіпотезу про високий ступінь не випадковості поєднання словоформ у межах незмінного компонента всіх проаналізованих моделей синтаксичних фразеологізмів. Водночас зменшення коефіцієнтів MI log Freq для конструкцій з низькою частотою є вмотивованим у програмах з автоматичного визначення колокацій / конструкцій, проте не є виправданим для вивчення синтаксичних фразеологізмів: вони можуть мати низьку частоту в певному корпусі текстів, але це не є достатнім доказом меншого ступеня зв'язаності відповідних словоформ.

Здійснений аналіз засвідчує, що показник асоціації MI є найбільш придатним для вірогідності встановлення стійкості поєднання двох або більше словоформ, що входять до складу незмінного компонента синтаксичного фразеологізму.

Перспективи подальших досліджень бачимо в залученні обчислення показників асоціації до створення комп'ютерної програми визначення синтаксичних фразеологізмів у тексті.

Таблиця 1

**Показники асоціації MI та MI log Freq для моделей синтаксичних фразеологізмів в українській мові за даними УНЛК**

з/п	Модель синтаксичного фразеологізму	Абсолютна частота вживання незмінного компонента синтаксичного фразеологізму	Абсолютна частота вживання словоформ, що входять до незмінного компонента синтаксичного фразеологізму	Показник асоціації MI	Показник асоціації MI log Freq
1	<i>Де (вже) там</i> N <sub>1</sub> /Adj/Adv Cop <sub>F</sub> /Inf	<i>де там</i> 552	<i>де</i> 4349 <i>там</i> 3326	12,75	78,92
		<i>де вже там</i> 59	<i>вже</i> 3876	28,4	113,99
2	<i>Не до</i> N <sub>2</sub> Cop <sub>F</sub>	<i>не до</i> 1841	<i>не</i> 4844 <i>до</i> 4845	13,79	101,62
3	<i>Ну і/ї</i> N <sub>1</sub> Cop <sub>F</sub>	<i>ну і</i> 823	<i>ну</i> 2423 <i>і</i> 4731	13,66	89,9
		<i>ну й</i> 897	<i>й</i> 4674	13,8	92
4	<i>Оце так</i> N <sub>1</sub> Cop <sub>F</sub>	<i>оце так</i> 400	<i>оце</i> 1548 <i>так</i> 4676	13,28	95,13
5	<i>Оце</i> N <sub>1</sub> <i>так</i> N <sub>1</sub> Cop <sub>F</sub>	<i>оце ... так</i> 70	<i>оце</i> 1548 <i>так</i> 4676	10,77	44,99
6	<i>Теж мені</i> N <sub>1</sub> Cop <sub>F</sub>	<i>теж мені</i> 201	<i>теж</i> 2680 <i>мені</i> 2582	12,35	64,29
7	<i>Чим не</i> N <sub>1</sub> Cop <sub>F</sub>	<i>чим не</i> 478	<i>чим</i> 2938 <i>не</i> 4844	12,56	76
8	<i>Що (ж це) за</i> N <sub>1</sub> Cop <sub>F</sub>	<i>що за</i> 2646	<i>що</i> 4843 <i>за</i> 4831	14,31	110,58
		<i>що це за</i> 738	<i>це</i> 4593	27,73	179,58
		<i>що ж це за</i> 206	<i>ж</i> 4240	41,27	296,11
9	<i>Яке (вже/ж) там</i> N <sub>1</sub> /Adj/ Adv	<i>яке там</i> 218	<i>яке</i> 3613 <i>там</i> 3326	11,67	61,67

	Сорр/Inf	<i>яке вже там 17</i>	<i>вже 3946</i>	23,47	66,509
		<i>яке ж там 7</i>	<i>ж 4240</i>	22,09	45,03
10	N <sub>1</sub> (він) і в Африці N <sub>1</sub> Сорр	<i>і в Африці 30</i>	<i>і 4731 в 4864 Африці 510</i>	26,3	88,56
		<i>він і в Африці 5</i>	<i>він 4009</i>	39,18	68,82
		<i>вона і в Африці 5</i>	<i>вона 4137</i>	39,13	68,74
		<i>воно і в Африці 0</i>	-	-	-
		<i>вони і в Африці 3</i>	<i>вони 4366</i>	38,32	52,08

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Величко А. В. Синтаксическая фразеология для русских и иностранцев : Учебное пособие / А. В. Величко. – М. : Изд-во МГУ, 1996. – 96 с.

2. Всеволодова М. В., Лим Су Ён. Принципы лингвистического описания синтаксических фразеологизмов: На материале синтаксических фразеологизмов со значением оценки / М. В. Всеволодова, Ён Лим Су. – М. : МАКС Пресс, 2002. – 164 с.

3. Залеская В. В. Программа выявления в тексте двучленных статистически значимых осмысленных коллокаций (на материале русского языка) / В. В. Залеская // Технологии информационного общества в науке, образовании и культуре : сборник научных статей. Труды XVII Всероссийской объединенной конференции «Интернет и современное общество» (IMS-2014), Санкт-Петербург, 19 – 20 ноября 2014 г. – СПб : Университет ИТМО, 2014. – С. 283–289.

4. Русская грамматика: В 2-х т. – Т. 2. Синтаксис / Под ред. Н. Ю. Шведовой. – М. : Наука, 1980. – 709 с.

5. Ситар Г. В. Статус синтаксичних фразеологізмів у системі фразеологічних одиниць / Г. В. Ситар // Вісник Донецького національного університету. Серія Б. Гуманітарні науки. – Донецьк : ДонНУ, 2011. – № 2. – С. 66–74.



6. Ситар Г. В. Статистичні критерії аналізу синтаксичних фразеологізмів / Г. В. Ситар // Вісник Донецького національного університету. Серія Б. Гуманітарні науки. – Вінниця: ДонНУ, 2015. – № 1-2. – у друці.

7. Хохлова М. В. Исследование лексико-синтаксической сочетаемости в русском языке с помощью статистических методов (на базе корпусов текстов) : автореф. дисс. на соискание ученой степени канд. филол. наук : спец. 10.02.21 «Прикладная и математическая лингвистика» / М. В. Хохлова / Санкт-Петербургский государственный университет. – Санкт-Петербург, 2010. – 26 с.

8. Ягунова Е. В., Пивоварова Л. М. От коллокаций к конструкциям / Е. В. Ягунова, Л. М. Пивоварова // АСТА LINGUISTICA PETROPOLITANA. Труды Института лингвистических исследований РАН. Т. X. Ч. 2. Русский язык: грамматика конструкций и лексико-семантические подходы / Ред. тома С. С. Сай, М. А. Овсянникова, С. А. Оскольская. – СПб. : Наука, 2014. – С. 568–617.

9. Church K., Hanks P. Word association norms, mutual information, and lexicography / K. Church, P. Hanks // Computational Linguistics. – #16(1). – 1990. – P. 22–29.

10. Everitt B. S. The Cambridge Dictionary of Statistics. 2nd edition / B.S. Everitt. – Cambridge : Cambridge University Press, 2002. – 410 pp.

11. Evert S. The Statistics of Word Cooccurrences: Word Pairs and Collocations / S. Evert : PhD dissertation, IMS, University of Stuttgart, 2004 (Published in 2005). – 353 P. – Free PDF available from <http://purl.org/stefan.evert/PUB/Evert2004phd.pdf>

12. Evert S., Krenn B. Using small random samples for the manual evaluation of statistical association measures / S. Evert, B. Krenn // Computer Speech & Language. – 2005. – 19(4). – P. 450–466.

13. Fano Robert M. Transmission of Information: A Statistical Theory of Communications / Robert M. Fano. – New York : The Technology Press, M.I.T., and John Wiley & Sons, Inc., 1961. – 389 pp.

14. Pecina Pavel. Lexical Association Measures. Collocation Extraction. Volume 4 of Studies in Computational and Theoretical

Linguistics / Pavel Pecina. – Prague: Institute of Formal and Applied Linguistics, 2009. – 143 pp.

15. Rychlě Pavel. A Lexicographer-Friendly Association Score / Pavel Rychlě // RASLAN 2008 Recent Advances in Slavonic Natural Language Processing. Second Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2008 Karlova Studbьnka, Czech Republic, December 5–7, 2008 Proceedings. – Brno, 2008. – P. 6–9: <https://nlp.fi.muni.cz/raslan/2008/papers/13.pdf>

16. Seretan V. Syntax-Based Collocation Extraction / V. Seretan // Text Speech and Language Technology. Series Editors Nancy Ide, Jean Vйronis. – Volume 44. – Dordrecht – Heidelberg – London – New York : Springer, 2011. – 222 pp.

17. Stubbs M. Collocations and semantic profiles: On the cause of the trouble with quantitative studies / M. Stubbs // Functions of Language. – 1995. – 2, 1. – Pp. 23–55.

СИТАРЬ А.

### **ВЫЧИСЛЕНИЕ МЕРЫ АССОЦИАЦИИ *MI LOG FREQ* КАК СТАТИСТИЧЕСКИЙ МЕТОД ИССЛЕДОВАНИЯ СИНТАКСИЧЕСКИХ ФРАЗЕОЛОГИЗМОВ**

Статья посвящена статистическому анализу синтаксических фразеологизмов на материале украинского языка. Выполнены вычисления меры ассоциации  $MI \log Freq$ , которая является модифицированной формулой коэффициента  $MI$  (mutual information) и дает возможность определить степень неслучайности (статистически доказанной связанности) сочетания словоформ, входящих в состав неизменяемого компонента модели предложения.

Представлены результаты выполненных вычислений для ряда моделей синтаксических фразеологизмов украинского языка по данным Украинского национального лингвистического корпуса.

*Ключевые слова:* конструкция, корпус текстов, синтаксический фразеологизм, статистический анализ, мера ассоциации, фразеологизированное предложение.

SYTAR H.

### **COMPUTATION OF *MI LOG FREQ* ASSOCIATION MEASURE AS A STATISTICAL METHOD OF SYNTACTIC IDIOMS RESEARCH**

The article is devoted to the statistical analysis of syntactic idioms based on the Ukrainian language. The computation of association measure of MI log Freq (mutual information logarithm frequency) was carried out, which is the modified formula of MI (mutual information) coefficient and allows us to determine the degree of randomness / non- randomness of word forms that are part of the permanent component of sentence model.

The results of the computations are performed for a number of syntactic idioms models of the Ukrainian language according to the Ukrainian National Linguistic Corpus. Use of a variant of the formula given in the work of Vera Zaleskaya is caused by the following factors: a) computer programs use natural logarithms, so the choice of this particular formula is driven by its further imposition to the corpus purposely to carry out the computations automatically; b) there was added 1 to the construction frequency to avoid program error if  $f(n,c) = 0$  (since  $\log 0$  does not exist).

Association measure of MI log Freq for all examined models of syntactic idioms is high, it is in the range from 44,99 (*оце ... так*) to 296,11 (*що ж це за*). However, unlike the MI coefficient, in case of computation of MI log Freq measure there was not registered the statistically significant link between the number of construction components and its value. Cases where MI log Freq value for the three-components constructions is lower than for the two-components (particularly, *яке ж там* 45,03 and *яке там* 61,67), or value for the four-components is lower than for the three-components (for example, *він і в Африці* 68,82, *вона і в Африці* 68,74, *вони і в Африці* 52,08 and *і в Африці* 88,56) are explained by the low frequency of constructions with a large number of components (respectively 7 and 208; 5, 5, 3 and 30).

The derived quantitative data according to formulas MI and MI log Freq confirm the advanced hypothesis of a high degree of word forms coherence within a permanent component of all the analyzed models of syntactic idioms. However, reduction of MI log Freq coefficients for constructions with low frequency is motivated in programs of automatic collocations / constructions detection, but it is not warranted for syntactic idioms research: they may have a low frequency in a particular corpus of texts, but this is not sufficient evidence of a smaller degree of coherence of corresponding word forms.

The conducted analysis shows that MI association measure is the most appropriate one for likelihood of establishing the stability of two or more word

forms combination that are part of the permanent component of syntactic idiom.

The results of computation of association measures can be involved into development of a computer program of syntactic idioms detection in the text.

*Key words:* construction, text corpus, syntactic idiom, statistical analysis, association measure, sentence with phraseological structure.

УДК 821.161.2 (045)

**СІБРУК Анастасія**

### **НАЗВИ КУЛЬТУВИХ РЕЧЕЙ ХРИСТИЯНСТВА, ЕЛЕМЕНТІВ ОЗДОЛЕННЯ ЦЕРКОВНИХ БУДІВЕЛЬ ТА ОПРАВ ІКОН У ДАВНЬОРУСЬКИХ ТЕКСТАХ XI – XIV ст.**

У статті розглядаються основні назви культових речей християнства та слова на позначення елементів оздоблення церковних приміщень у східнослов'янській писемності XI–XIV ст., їх історія, етимологія та лексичне значення.

**Ключові слова:** Київська Русь; лексеми *кръсть*, *тһльникъ*, *цата*, *трһсна*; східнослов'янська писемність XI–XIV ст.

На сьогоднішній день в україністиці постає **проблема** створення ґрунтовної історичної лексикології української мови. Вивчення східнослов'янського лексичного фонду XI–XIV ст., реконструкція семантичної структури окремих іменників надзвичайно важливі для простеження історії давньоруської мови. **Актуальність роботи** зумовлена тим, що словниковий склад української мови в історико-лінгвістичному аспекті описаний недостатньо. Питання історії як окремих слів, так і лексико-семантичних груп привертало та привертає увагу лінгвістів. **Мета статті** подати лексико-семантичний аналіз назв культових речей християнства за **матеріалами** писемних східнослов'янських пам'яток XI–XIV ст. Впровадження християнства на Русі сприяло розвитку прикладного мистецтва, що призвело до масового виробництва речей християнського культу – хрестів, іконок,