

## КОМП'ЮТЕРНО-ЛІНГВІСТИЧНІ МЕТОДИ ПЕРЕВІРКИ ДОСТОВІРНОСТІ ПЕРСОНАЛЬНИХ ДАНИХ КОРИСТУВАЧІВ ВЕБ-СПІЛЬНОТ

© Пелешишин А.М., Федушко С.С., 2010

Розглянута актуальна проблема дослідження та розроблення методів перевірки достовірності наданих особистих даних користувачів веб-форуму. Досліджено методи гендерних та вікових відмінностей мови написання повідомлень користувачів веб-форуму.

**Ключові слова:** методи перевірки достовірності, гендерні відмінності, вікові відмінності, веб-форум, користувачі, аккаунт.

**This article considers current problem of investigation and development of authentication of the given Web-forum member's personal data. Methods of gender and age differences of Web-forum member's messages are investigated.**

**Keywords:** authentication methods, gender differences, age differences, Web-forum, members, account.

### 1. Постановка проблеми

Веб-спільноти акумулювали в собі величезну базу контактів та профілів, в яких міститься чимало інформації про людину та її уподобання. Контингент веб-спільнот – надзвичайно велика кількість людей незалежно від віку, статі, професії, освіти, національності, сімейного стану тощо, яким для реєстрації необхідно заповнити форму з персональними даними. Але постає питання: чи достовірні ці дані, які подають учасники веб-спільноти для реєстрації? І саме лінгвістичне дослідження відмінностей стилів написання повідомлень користувачами є основою для розроблення ефективних методів перевірки опублікованої інформації у аккаунтах користувачів. Отже, важливо допомогти адміністратору та модератору у моніторингу і перевірці достовірності наданих користувачами веб-форуму персональних даних.

Дослідження спрямоване на розроблення комп'ютерно-лінгвістичних методів перевірки достовірності персональних даних користувачів веб-спільнот. На основі аналізу інформаційного наповнення чотирьох веб-спільнот (“Львів. Форум Рідного Міста”, “Тернопільський Форум”, “Родинний форум “Малеча”, “ROCK.LVIV.UA: Форум”) ми дослідили комп'ютерно-лінгвістичні аспекти перевірки достовірності наданої користувачем інформації, основані на класифікації категорій стилів написання повідомлень користувачами веб-спільнот.

Розроблення комп'ютерно-лінгвістичних методів перевірки достовірності наданої інформації користувачем допоможе вирішити ряд актуальних проблем:

– регулярна поява великої кількості *псевдоанкет*, особливо псевдоанкет відомих людей. Важким завданням є виявлення справжньої анкети користувача, якщо таких анкет є декілька й інформація, яка міститься в них, більш-менш правдоподібна;

– *фішинг* [19] – використання інтернет-технологій з метою отримання доступу до конфіденційної інформації про користувачів (паролів, логінів, фінансової інформації тощо). Найчастіше жертвою фішингу стає дитина. Діти, як правило, не замислюються, для чого стороннім потрібна така інформація;

– *кібербулінг* [19] – переслідування інших осіб з використанням Інтернету (електронної пошти, форумів, чатів тощо). Яскравим проявом кібербулінгу є видавання себе за іншу особу в текстових повідомленнях з нецензурними висловлюваннями та образами;

– подання користувачами веб-спільнот в свої аккаунти неправдивої інформації.

Актуальність цього дослідження зумовлена зростанням кількості випадків, коли користувачі стають жертвами кібербулінгу, фішингу, не підозрюючи, що спілкуються не з однолітком чи другом, а із зовсім чужою людиною. Вирішення цих проблем зумовлює потребу саме розроблення комп'ютерно-лінгвістичних методів перевірки достовірності наданої користувачем інформації, які полягають у визначенні категорій стилів написання повідомлень користувачів веб-спільнот.

## 2. Аналіз останніх досліджень

Аналіз сучасних праць науковців виявив, що, незважаючи на велику кількість досліджень мови, вагомий внесок зробили вчені саме в дослідження гендерних та вікових відмінностей в мові. Ці відмінності в мові проаналізовано в наукових працях видатних західноєвропейських та американських науковців. У їхніх наукових роботах доведено та досліджено наявність гендерних та вікових відмінностей в англійській мові. Предметом нашого дослідження стали східнослов'янські мови, а саме українська.

Результати дослідження гендерних та вікових відмінностей в українській мові є основою для розроблення комп'ютерно-лінгвістичних методів перевірки достовірності наданих користувачами веб-спільнот Укрнету персональних даних.

У значній кількості досліджень відмінностей в мовах, які використовують чоловіки і жінки, розглянуто багато наукових теорій. Вчені Мулек, Брадак і Гібсон [12], узагальнивши більше ніж 30 досліджень, знайшли 16 мовних особливостей, які гендерно диференціюють власників повідомлень: п'ять досліджень показали, що чоловіки частіше віддають перевагу *кількісній оцінці* ("передбачається збільшення прибутків на 13 %"), ніж жінки, а результатом трьох досліджень є те, що чоловіки часто використовують *прикметники, які описують оцінку стану* ("Це можна продовжити завтра"). З іншого боку, шість досліджень показали, що жінки частіше вживають *підсилювальні прислівники* ("Твоя робота насправді є важливою") та п'ять досліджень – *фрази емоційного характеру* ("Якщо б він любив те, що він робить ..."). Хоча такі відмінності мови відзначаються часто, їх не слід розглядати як остаточний результат. Незважаючи на це, науковці вважають, що гендерні відмінності в мові потрібно розглядати в контексті комунікації.

Робін Лакофф [9] досліджувала гендерні відмінності на рівні конкретних фраз. Вона визначила в мові жінок два типи конкретних фраз: ухилення від відповіді (наприклад, "Здається, що...") і питання (наприклад, "..., правда?", "хіба не так?"). Використання жінками фраз ухилення від відповіді інтерпретується як менша схильність їх до нав'язування своїх думок іншій особі. Згідно з цією ідеєю Лакофф [9] стверджувала, що жінки частіше, ніж чоловіки в тій самій ситуації, висловлюються в екстрав'єрсивній формі (напр., "Не могли б ви..").

Томсон і Мурачвер [17], вивчаючи електронну комунікацію, виявили, що чоловіки і жінки однаково часто задають питання; висловлюють компліменти, вибачення, думки, образи інтернет-товаришам. Інші дослідження [10] повідомляють про істотні відмінності з цим результатом. Як показало порівняння мови 36 жінок і 50 чоловіків-менеджерів, чоловіки використовують значно більше заперечень і задають запитань, а жінки – вказівок. Проте дослідження не підтверджує, що чоловіки використовують загалом більше слів, тоді як жінки – довші речення. Одним з можливих пояснень цих суперечливих відомостей є те, що різні контексти, з яких були отримані зразки повідомлень, вплинули на розмір та напрями гендерних відмінностей. Згідно з попередніми дослідженнями [1], мова жінки складніша, заплутаніша, ніж мова чоловіків.

Деякі дослідження доводять [11] використання розділових речень жінками, хоча вчені Дубоїс і Кроуч [6] дійшли протилежних висновків. Вони надали додаткові докази, що жінки викорис-

тують фрази, які часто означають почуття невизначеності, вагання, мінливості. Частіше в письмових повідомленнях жінок виявляють фрази, в яких перцептивні або когнітивні дієслова поєднуються з займенниками першої особи однини, наприклад, "Цікаво, що (котрий)...". Однак стиль мовлення чоловіків є конкурентоспроможнішим в контексті онлайн-дискусій [8].

Гендерні відмінності були також розглянуті у дискусіях між користувачами мережі. Дослідження стилів мови в онлайні показало [10], що характеристики стилів мови чоловіків і жінок в інтернет-спілкуванні порівняно з реальною розмовою істотно не змінюються. Жінки використовують більшу кількість підсилювальних прикметників, сполучників (напр., "але") й допоміжних модальних дієслів (напр., мати можливість) і багато запитань. Чоловіки вживають більше клятв, присяг, довші слова та використовують додаткові посилання на місцезнаходження і соціальний статус [11].

Традиційні гендерні дослідження мови спрощують її структури в онлайн-спілкуванні [14], тому що в онлайні більше свободи і гнучкості.

Під час дослідження 3000 онлайн-повідомлень Вітмер і Кацман [20] виявили, що жінки частіше використовують графічні акценти, зокрема смайлики, щоб висловити свої емоції в дискусії.

Георг Кейз і Джон Шутлеворт [15] заявили, що:

- *жінки* використовують більше слів для висловлення думок, розмовляють ввічливіше, нерішучіше / невпевненіше, скаржаться і прискіпуються, задають більше питань, підтримують одне одного;
- *чоловіки* більше присягаються, не говорять про емоції, а радше про спорт, роботу, жінок і машини, є конкурентоспроможними і домінують у розмові, говорять авторитетніше, використовують вказівки та інструкції у спілкуванні.

Мовознавці все більше доходять до консенсусу [8], що чоловіки використовують мову передусім з важливою метою передавання інформації; жінки більш схильні використовувати спілкування, цілеспрямоване в соціальному напрямі.

Отже, гендерний фактор у мовленнєвому стилі потрібно досліджувати, як правило, не поодиноці, а спільно з іншими соціальними факторами, такими як вік, клас, етнічна приналежність, місце проживання тощо.

Недавні дослідження [12] дають змогу припустити, що використання мови також варіюється залежно від віку людини, а також що гендерні відмінності є іншими для дітей різного віку. Тому ми проаналізували ознаки мови підлітків.

Одна з ознак написання повідомлень підлітками – використання великої кількості графічних усмішок (смайликів) у тексті повідомлення. Скотт Фалман, професор психології університету Карнегі-Меллона у Пітсбурзі, надіслав колегам електронне повідомлення, у якому запропонував використовувати смайлик (smiley, emoticon) для вираження емоцій у інтернет-комунікаціях. Професор запропонував використовувати для позначення жартівливих слів символ – :-), щоб позначити сумну звістку – :-(.

*Смайлики* – графічні символи або знаки пунктуації, за допомогою яких виражають емоції. Смайлики ввели через відсутність можливості бачити міміку і жести тіла в онлайні. Вони слугують для вираження думок та заповнення порожнечі в онлайн-комунікації [5]. Також є інший стиль смайликів [19] – "каомодзі" (азіатські смайлики). На відміну від "класичних" смайликів, їх не треба перевертати при читанні, головна увага приділяється очам, а не роту: (^\_^) – посмішка, (^\_~) – моргання, (<\_>) – сум, (O\_o) – здивування тощо.

Результатом наступних досліджень [18] є протилежне твердження – наявність смайликів затьмарює словесний зміст у тлумаченні повідомлень. Але вони допомагають прослідкувати відносини між користувачами мережі. Наприклад, у дослідженні дискусій чату [2] користувачі, які використовують смайлики, сприймаються як більш "динамічні", "дружні", "цінні", і "балакучі", ніж користувачі, які не застосовують цих символів.

Отже, у використанні смайликів також існують вікові та гендерні відмінності. Дослідивши 3000 онлайн-повідомлень, Діана Вітмер [21] виявила, що жінки використовують більше графічних акцентів, зокрема смайлики, щоб висловити свої емоції в дискусії, ніж чоловіки. Гендерні відмінності також є у виборі користувачами аватарів та інтернет-імен. *Інтернет-ім'я* (нікнейм) впливає на соціальну взаємодію. Позитивні прізвиська сприяють швидкій адаптації в середовищі однолітків [4]. *Аватари* (аватарки, ави) – це різноманітні картинки, фото, які відображають характер людини, її світогляд, адже зображення на аватарі користувач вибирає сам. Аватар завжди розташований поруч з інтернет-іменем.

Проаналізувавши виконані дослідження, ми зробили висновок, що вже виявлено відмінності у мові соціальних груп, проте ці відмінності недостатньо досліджені. Детальніше дослідження відмінностей стилів мовлення допоможе визначити вік, стать, національність, професію та інші особисті характеристики, тобто особистість зареєстрованих користувачів Мережі.

### 3. Цілі статті

Мета цього дослідження – розроблення комп'ютерно-лінгвістичних методів для перевірки достовірності персональних даних, наданих користувачами на основі дописів на веб-форумі.

Методи перевірки достовірності особистої інформації, яку надав користувач, полягають у дослідженні відмінностей стилів мовлення користувачів в інтернет-дискусіях. Ми дослідили гендерні, вікові, соціолектичні, діалектичні відмінності стилів написання повідомлень за допомогою методів квантативного мовознавства. Наше дослідження ґрунтується на класифікації категорій стилів написання повідомлень користувачами веб-спільнот.

Отже, цілями цієї статті є:

- аналіз наявних досліджень відмінностей стилів написання повідомлень окремих соціальних груп;
- визначення рівня розкриття особистої інформації в аккаунтах користувачів, наприклад, таких даних, як ім'я, вік, стать, освіта або адреса;
- дослідження особливостей поведінки підлітків у мережі;
- встановлення основних характеристик й особливостей використання молодіжного сленгу, соціальних діалектизмів, безеквівалентної лексики у інтернет-комунікації, за допомогою аналізу інформаційного наповнення веб-спільнот;
- визначення критеріїв ідентифікації віку, статі, національності, професії користувачів у веб-спільнотах (“Львів. Форум Рідного Міста”, “Тернопільський Форум”, “Родинний форум “Малеча”, “ROCK.LVIV.UA: Форум”), методом аналізу мови написання повідомлень та за типом інтернет-імені, аватарів, графічних засобів вираження емоцій;
- дослідження гендерних особливостей написання повідомлень підлітків і дорослих користувачів мережі, на основі інформаційного наповнення веб-спільноти “Форум Рідного Міста”;
- визначення ймовірності ідентифікації авторів повідомлень в інтернет-спілкуванні.

### 4. Основний матеріал

#### 4.1. Розкриття особистої інформації в Інтернеті

У цьому статистичному дослідженні розкриття особистої інформації в Інтернеті проаналізовано персональну інформацію, яку вказали користувачі у своїх аккаунтах. А саме: нікнейм, повне ім'я, вік, дата народження, місце проживання, освіта та контактна інформація (адреса електронної пошти, посилання на інші особисті веб-сторінки). Ми проаналізували дані аккаунтів користувачів чотирьох веб-спільнот (“Львів. Форум Рідного Міста”, “Тернопільський Форум”, “Родинний форум “Малеча”, “ROCK.LVIV.UA: Форум”).

Результати нашого гендерного дослідження надання користувачами веб-спільнот особистої інформації в свої аккаунти подано тна рис. 1, 2.

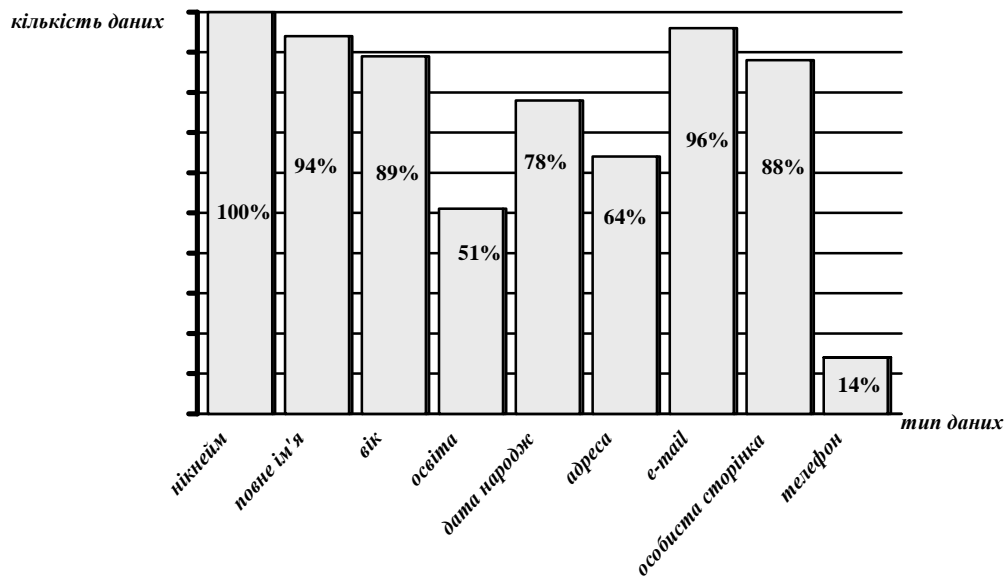


Рис. 1. Рівень надання персональної інформації чоловіками

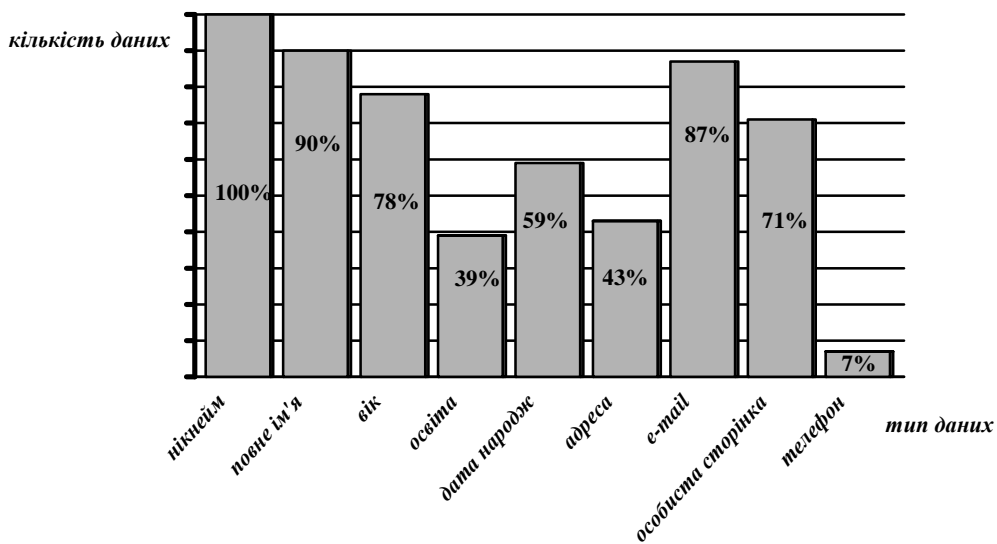


Рис. 2. Рівень надання персональної інформації жінками

Отже, дослідження доводить, що чоловіки більшою мірою, ніж жінки надають персональну інформацію про себе у аккаунтах: повне ім'я (на 4 %), вік (на 11 %), освіта (на 12 %), дата народження (на 19 %), адреса (на 21 %), а також контактну інформацію: e-mail (на 9 %), особиста сторінка (на 1 %), телефон (на 7 %).

## 4.2. Особливості поведінки підлітків у Мережі

Часто підлітки реєструються у приватних чатах, форумах та спілкуються на будь-які теми, видаючи себе за дорослих. Хлопці в підлітковому віці віддають перевагу всьому, що виходить за межі дозволеного: брутальний гумор, насильство, азартні ігри, еротичні та порносайти. Дівчата схильні до фривольних розмов, видають себе за дорослих жінок, внаслідок чого стають жертвами сексуальних домагань. Також часто вони розміщують на своїх сторінках провокаційні фото.

Ми виконали опитування 153 дітей віком від 6 до 18 років за допомогою анкетування. Так, понад 28 % опитаних дітей готові переслати свої фотографії незнайомцям у веб-мережі. 17 % без коливань погоджуються повідомити інформацію про себе і свою сім'ю – місце проживання, професію і графік роботи батьків, наявність в будинку цінних речей тощо. Діти, як правило, не замислюються, для чого стороннім потрібна така інформація.

Загалом про небезпеку інтернет-сайтів знають тільки від 3 до 12 % опитаних нами дітей. Чотирнадцятирічна дівчинка у своїй анкеті описала таку ситуацію: однокласниця познайомилась через ICQ із ровесником, а на зустріч прийшов дорослий тридцятирічний чоловік.

Дослідивши поведінку підлітків і дітей в Інтернеті, ми зробили висновок, що діти та їх батьки недостатньо проінформовані про небезпеку веб-ресурсів, якими користується дитина. Насторожує також готовність дитини надати будь-яку особисту інформацію незнайомій людині.

## 4.3. Вікові та гендерні відмінностей у мові написання повідомлень

### 4.3.1. Соціолект підлітків

Соціолект підлітків формується під впливом соціокультурних факторів (кіно- та шоу-бізнесу, ЗМІ, новітніх технологій соціальної комунікації, музичної субкультури, сленгу афроамериканців тощо) на базі соціально-психологічної спільності дітей у підлітковому віці (протиставлення себе дорослим, домінування однопітків у колі співрозмовників, надмірна емоційність, підвищена агресивність тощо) та швидко реагує на різні зміни соціального характеру.

З сучасного українського молодіжного жаргону можна навести приклади таких синонімічних рядів [22]: *торба – гаплик – хана – труба – глина – слива – ландиш (кінець, крах чогось), герла – матильда – клава – тітка (дівчина)* тощо.

Проаналізувавши інформаційне наповнення молодіжних форумів ( “ Тернопільський Форум”, “ROCK.LVIV.UA : Форум”), ми сформували характерні риси стилю мовлення підлітків:

1. Позначення вигуків (огооооо..., Аучччч!!!).
2. Використання скорочень “треба, хочеш” скорочують на “тре, хош”, “дякую” – “дякс, дякі”.
3. Написання перших літер слів у словосполученні “має бути” пишуть “мб”, “хтозна” пишуть “хз”.
4. Транслітерація слів російської та англійської мов (“ван лов, ван хард, летс гет тугезер енд філ олрайт”), “Ху а зіз піпл???”, “канцерт в амерікі”).
5. Вираження сміху – “ггггг)))).”.
6. Велика кількість графічних усмішок :-)) ;
7. Використання послідовності дужок ))))))) замість графічної усмішки :-).
8. Абrevіатурні скорочення (імхо – на мою скромну думку).
9. Специфічна лексика з комп’ютерних ігор (ламер, геймер).
10. Використання олбанської мови – стилістичного прийому, спрямованого на досягнення експресивно-комічного ефекту шляхом зумисно помилкового написання слів, відмова від пунктуації (lol – lutz, “тупа зачод!”, “мальчіГ”).
11. Використання цифр як літер (то4но, клб080).
12. Застосування неалфавітних знаків ( ! замінюють літеру І – !нод! ).
13. Написання всіх літер, окрім голосних, великими буквами (“Це ТоЧно Не ФаРТ”).
14. Помилки найчастіше не виправляються.
15. Неалфавітні й нецифрові символи можуть складатися у літери (комбінації косих рисок  $\wedge$  – літера М,  $|<$  – літера К,  $[-$  – літера Е)).

16. Надмірне вживання слів ненормативної лексики з пропуском букв чи їх заміною на інші букви.

17. Використання фамільярної лексики (“дружбан”, “чуділо”, “мурло”).

18. Вживання вульгаризмів (“морда”, “пика”).

Здійснене дослідження доводить, що соціолект підлітків – це некодифікована норма мовного спілкування підлітків, яка визначається використанням нестандартної лексичної системи, запозичень зі соціолектів інших молодіжних субкультур, власних лексичних новоутворень, певних синтаксичних конструкцій, морфологічних форм, графічних позначень, неалфавітних символів тощо.

#### 4.3.2. Критерії вибору інтернет-імен та аватарів

Користувачі веб-спільнот приділяють велику увагу своїм інтернет-іменам та аватарам, адже вони є “віртуальним альтер его” користувача, підкреслюють його індивідуальність. Інтернет-імена та аватари можуть інколи сказати більше про людину, її захоплення, характер, ніж інформація, опублікована у аккаунті користувача. У цьому дослідженні ми звернули увагу також на методи та критерії вибору імен та аватарів користувачів веб-спільнот (“Львів. Форум Рідного Міста”, “Тернопільський Форум”, “Родинний форум “Малеча”, “ROCK.LVIV.UA: Форум”).

Проаналізувавши інтернет-імена, які вибрали для реєстрації та онлайн-спілкування учасники веб-спільнот, ми сформуваємо такі критерії вибору інтернет-імен користувачами веб-форумів. Розділили їх на типи:

1) імена людей (*Шанак Тетяна, Олена, Sonja*);

2) імена, пов'язані з політичними поглядами (*націоналіст, комуніст*);

3) імена, пов'язані з назвами природних явищ, рослин, тварини тощо (*Сова, Форест, fish, Вітер, осінь, зіронька, Panda, Crobak, root*);

4) імена, пов'язані з територіальною приналежністю (*Східняк, львів'янин, Московит, Moscal*);

5) використання художніх засобів мови (світлий вечір, крижана посмішка, море сліз);

6) імена, пов'язані з літературою, міфологією, фільмами, історією і відомими людьми (*Seneka, Kallisto, Відьмочка, чарівник з міста Ромельн, козак*);

7) імена сексуального характеру (*serceedka, LLLover*);

8) імена, які описують характер, емоційний стан (*сумна, freewill, ReadHead, freak, пішучий, бравий*);

9) імена, пов'язані з професією (*Соло, репортер, каунтер, artist*).

Цілоком інші результати у дослідженні інтернет-імен, якими користуються діти:

1) імена (*Сашко, Олег, Katrusja*);

2) прізвиська (*21dimas, MarkX, Fredi*);

3) імена-назви тварин (*хом'як, bunny, Wild-cat, Pig*);

4) імена, пов'язані з інструментами або виявами насилля (*калашнікоф, меч, breather, breaker*);

5) імена, які описують емоції, особисті характеристики (*самотність, бліда поганка, іржа, зараза, abnormal, скромна*);

6) імена, пов'язані з музикою та поп-культурою (*Kurt, sound, Hardmetalgirl*);

7) імена, пов'язані з фантазією/міфологією, фільмами, казками (*дуть місяця, Andriy Haus, незнайка, alien, femida, drakyla, русалка, PoKEMoN*);

8) імена, які допомагають самоствердитись (*sexygirl, superman, egoman, Coolgirl*);

9) лагідні і пестливі імена (*Lovely, kotenok, Angelbaby*);

10) імена, символізують природні явища (*зіронька, веселка, крижана зірка, sun*);

11) безглузді імена (*уть ють ать ять, трамбулямбулямбус, кілька в томаті*).

Результатом дослідження використання інтернет-імен дорослих користувачів, є те, що, як правило, користувачі для реєстрації частіше подають свої справжні імена і прізвища порівняно з підлітками.

У дослідженні ми виявили гендерну і вікову закономірності вибору інтернет-імен користувачами чотирьох веб-спільнот (“Львів. Форум Рідного Міста”, “Тернопільський Форум”, “Родинний форум “Малеча”, “ROCK.LVIV.UA: Форум”).

Жінки віддають перевагу іменам, які описують характер, емоційний стану; пов'язані з літературою, фільмами і відомими людьми; використанню художніх засобів мов у формулюванні

імені. А чоловіки – іменам, пов'язаним з професією, такими, які вказують на політичні погляди, та імена, пов'язані з містом, країною. З однаковою частотою як чоловіки, так і жінки вибирають імена, пов'язані з літературою, міфологією, фільмами, історією і відомими людьми, та імена, пов'язані з назвою природних явищ, рослин, тварин тощо.

Хлопчики, як правило, вибирають імена, пов'язані з насиллям, з музикою та поп-культурою, прізвиська, а дівчатка – імена, які символізують природні явища, лагідні та пестливі імена, імена, які описують емоції, особисті характеристики. Безглузді імена, імена, які пов'язані з фантазією/міфологією, фільмами, казками, назвами тварин, допомагають поліпшити самооцінку, символізують природні явища, діти вибирають незалежно від статі.

Також ми дослідили гендерну закономірність *вибору аватарів підлітками*. Ми виявили, що дівчатка переважно вибирають на аватарки картини тварин, персонажів з мультфільмів, фото зі своїм другом, музичних кумирів, привабливих жінок або свої спокусливі фото. А ось хлопчики – фото відомих спортсменів, кіноакторів, музичних кумирів, сексуальних дівчат, картини, зображення на яких вирізняється жорстокістю.

Отже, інтернет-імена та аватари надають первинної форми онлайн-ідентичності і дають можливість ідентифікувати стать і вік користувача.

#### 4.3.3. Графічні засоби вираження емоцій – графічні усмішки (смайлики)

Основними графічними засобами вираження емоцій є графічні усмішки (смайлики). Смайлик – це письмове вираження емоцій у текстових повідомленнях символами ASCII-графіки. Часто у веб-спільнотах варіанти текстових усмішок автоматично перетворюються на графічні.

Широкої популярності у форумах, чатах набули графічні усмішки завдяки зручності вставлення у текст, зрозумілості, можливості вираження та уточнення закладених у повідомленні відправника емоцій текстом або графічним значком.

Використання графічних і текстових усмішок користувачами у діалогах проаналізовано у цьому дослідженні. Ми підраховали кількість використаних представниками обох статей смайликів та поділили найвживаніші смайлики на типи (табл. 1).

Таблиця 1

Використання графічних усмішок залежно від статі користувача

Тип смайлика	Графічне позначення	F	M
радість	: ) чи :-)	44%	23%
сум	: ( чи :-(	17%	31%
злість	>:( чи :O	8%	13%
кокетливість	; ) чи :P, :-*	21%	27%
втома	n/a	6%	1%
інші	:-/ чи :-  та ін.	4%	5%

Оскільки жінки традиційно емоційніші, то ми очікували, що жінки у своїх повідомленнях використовуватимуть смайлики частіше, ніж чоловіки.

Проте, всупереч прогнозам, виявлено тенденцію, що чоловіки використовують більше смайликів, ніж жінки (F=44 %, M=56 %).

Причому чоловіки у своїх повідомленнях використовували більше *кокетливих* та *сумних* смайликів, а жінки – *радісних*.

Отже, графічні усмішки – альтернативний шлях вираження індивідуальності для інтернет-користувачів. Дослідження використання смайликів дає можливість визначити гендерну приналежність автора повідомлення.

#### 4.3.4. Аналіз гендерних відмінностей мови інтернет-користувачів на основі дописів на веб-форумі “Львів. Форум Рідного Міста”

Проаналізувавши інформаційне наповнення “Львів. Форум Рідного Міста”, ми сформуваємо перелік базових ознак стилю написання повідомлень представниками як чоловічої, так і жіночої статі.



## **Критерії, які використовують жінки в інтернет-комунікаціях (визначені на основі інформаційного наповнення веб-форуму)**

### **На рівні речень :**

- підрядні речення обставини – відповіді на запитання: *як?, коли?, де?* головного речення. ("Будь там, де ...."; "Тому що на деревах є ще сніг ..., то виглядає ніби ..."; "Йди туди, звідки прийшов");
- підрядні речення, позбавлені змісту ("Ти знаєш, ...");
- підрядне речення, яке слугує для уточнення ("... , яка в основному вкрита ..."; "...там, де зустрічались раніше "; " ..., в якому щось ...");
- заперечне речення, твердження відсутності чогось ("Ви не відчуваєте, як майбутнє...");
- запитання – вказівки в питальній формі не беремо до уваги ("Що таке ..?" "Що ви робите?");
- вживання розділових речень типу: "Ви збираєтесь пообідати, чи не так?";
- прямі цитування ("Вона мені просто сказала: "Діти – це Чудо і вони не приходять на "вимогу."");
- протиставні речення ("Сніг зараз падає, але він скоро розтане ");
- окличні речення ("Боже, та звідки я знаю!", "Приєднуюся до привітань!");
- вживання речень, які висловлюють підтримку ("Не переймайся так, все буде добре", "Молодці, справились з цим");
- використання непрямих команд і прохань: ("Тобі не здається, що тут холодно?" – насправді це прохання увімкнути опалення чи зачинити вікно).

### **На рівні частин мови, конкретних слів та фраз:**

- згадування минулих подій, які викликають негативні емоції ("...як він міг так підставити?").
- використання дієслів умовного способу: ("Якби я могла б, то ...");
- велика кількість вибачень: ("Мені дуже шкода, але я думаю, що ..."; "...вибачте!");
- згадка про емоції та почуття ("щасливий", "привабливою", "пригнічує");
- ухиляння від відповіді; вживання слів, які вказують на відсутність довіри чи впевненості у достовірності висловлювання ("ніби", "на киталт", "вірогідно", "щось на зразок", "можливо", "мабуть");
- дієслівні фрази, які вказує на брак впевненості ("Цікаво, якщо б ...", "Здається ...", "Я не впевнена ...", "Ми можемо помилятися..");
- прикметники, які вказують на особисту оцінку, а не на опис почуттів, речей тощо ("бентежні", "неспокійний", "милий ...");
- акцент на словах – "отже", "дуже", "досить";
- використання 'порожніх' прикметників (прикметники, як не містять ніякої вагомості інформації): "божественна", "милі", "чарівні" тощо;
- вживання спеціальної лексики: жінки використовують більше таких слів, які позначають соціальні, родинні теми дискусій (чай на вагу, посуд з тефлоновим покриттям, сімейне життя);
- використання модальних слів: ("може", "буде", "мабуть", "повинно" – "Повинні ми бути там? Зробити це до обіду?");
- уникнення вульгарних і лайливих висловів ("не буду я негідно виражатись, але...");
- вживання підсилювальних слів: "особливо", "настільки", "надзвичайно", "справді", "досить", "огидно" ("Я надзвичайно рада, що ти прийшов!");
- використання особового займенника "ми" ("Ми вирішили не йти туди, щоб..");
- вживання пестливих слів: слів, які виражають ласку і ніжне ставлення, приязнь ("будь ласочка", "котик");
- використання ввічливих словоформ: "Ви не заперечуєте ...", "Я була би дуже вдячною, якби...", "... якщо ви не заперечуєте", "Щиро вдячна вам...", "Дякую за..").

## **Критерії, які використовують чоловіки в інтернет-комунікаціях (визначені на основі інформаційного наповнення веб-форуму)**

### **На рівні речень :**

- еліптичні речення – відсутність дієслівного присудка, що не встановлюється з контексту чи ситуації. ("Воно так все складно", "Як добре зроблено");

- речення, які слугують для вираження згоди ("*Ти правий!*", "*Я згідний з вами*", "*Так*", "*Ок!*");
- перефразування речень ("*Згідно з твоїми словами, ти став виправним циніком*");
- вказівки ( "*Обдумайте цю тему*", "*Чому це не покладете на місце?*");
- позначення вигуків ("*Ух*", "*Хм...*", "*Оце так!*");
- безособові речення ("*Розумію*", "*Погано*");
- обговорення поточних проблем ("*на сьогоднішній час ціни...*", "*нині виникла потреба...*", "*зараз існують не дуже якісні безплатні аналоги платних програм*");
- згадування минулих подій відносно теперішнього часу ("*...ще коли працював на старій роботі..*", "*...за годину тиждень тому обійшов!*", "*у 2003 було дешевше порівняно з аероклубом*").

#### ***На рівні частин мови, конкретних слів та фраз:***

- використання спеціальної лексики: розмови на теми спорту, технології, роботи, політики, грошей ("*а сплатити кредит треба негайно*", "*про керівництво держави мова йде*", "*бензинових криз*", "*матч*");
- посилення на кількість, величину, будь-яка згадка суми, кількості. ("*нижче за 28 ° C*", "*більшої частини регіону*", "*6-8 тисяч метрів висота*", "*на 10 см вищий за...*");
- вказівка на розташування об'єктів ( "*поруч ...*", "*на задньому плані*");
- використання займенника першої особи однини : я, мене і моїх ("*Я думаю...*"; *я повинен...*");
- географічні посилення ("*в нашому місті*", "*в цій країні*", "*ще далеченько їхати на захід від нашого міста*" );
- просторові посилення ("*там, біля магазину, який розташований на цій вулиці*");
- використання стверджувальних фраз ("*Я впевнений...*", "*Я точно знаю..*");
- вживання слів-розпоряджень ("*зроби...*", "*подивись на ...*");
- використання теперішнього часу ("*сьогодні*", "*зараз*");
- вживання довгих слів ("*конструктивний*", "*ідентифікував*", "*позачерговою*", "*переобладнаний*");
- часте використання присяг і клятв ("*Присягаюсь тобі*", "*..., чесне слово!*");
- вказівні займенники – той, ті тощо ("*в той день були змагання*", "*тоді я був не вдома*");
- використання прийменників, які вказують на час, місце розташування ("*з початку місяця*", "*там, збоку...*", "*згідно з*");
- використання лайливих і грубих слів ("*що ти верзеш?*", "*дурнуватий вчинок*");

Результати наших досліджень гендерних відмінностей у мові користувачів веб-форумів наведено у *табл. 2*. Також ми обчислили ймовірність появи кожної сформованої нами гендерних категорій стилю мови у дискусіях досліджуваних веб-форумів.

Порівнюючи характерні риси використання мови чоловіків і жінок у дискусіях веб-форуму "Львів. Форум Рідного Міста", загалом ми дійшли висновку, що є невеликі, але послідовні гендерні відмінності у мові під час онлайн-спілкування.

Отже, чоловіки в написанні повідомлень частіше використовують *вказівні займенники* ( $P=0,042$ ), а особливо *займенник першої особи однини – я, мене, моїх та ін.* ( $P=0,095$ ). Проте характерна ознака мови жінки – *використання займенника "ми"* ( $P=0,025$ ).

*Слова підтримки* ( $P=0,0397$ ), *слова, які виражають емоції та почуття* ( $P=0,036$ ), *ухилення від конкретної відповіді* ( $P=0,029$ ), *вибачення* ( $P=0,024$ ), *підсилювальні слова* ( $P=0,0224$ ) та *пестливі слова* ( $P=0,014$ ) характерніші для мови жінок.

Також жінки використовують більше слів, пов'язаних з психологічними та соціальними процесами, *згадування про дім, дітей, магазини* ( $P=0,0306$ ), тоді як чоловіки дискутують на *теми спорту, політики, роботи та грошей* ( $P=0,084$ ).

Результати цього дослідження підтверджують раніше зроблені висновки вчених, що чоловіки використовують більше *образливих форм* ( $P=0,041$ ) та *присяг і клятв* ( $P=0,049$ ), тоді як жінки – *ввічливі словоформи* ( $P=0,052$ ), *непрямі команди і прохання* ( $P=0,0098$ ) та *уникають вульгарних і лайливих висловів* ( $P=0,008$ ).

Таблиця 2.

## Критерії, які використовують користувачі в інтернет-комунікаціях

Тип аналізу	Характерні категорії написання повідомлення	PF	PM	
На рівні частин мови, конкретних слів та фраз	Велика кількість вибачень	<b>0,024</b>	0,0074	
	Модальні конструкції	<b>0,0239</b>	0,0088	
	Вказівні займенники – той, ті, того тощо	0,025	<b>0,042</b>	
	Згадка про емоції та почуття	<b>0,036</b>	0,0013	
	Просторові посилання	0,003	<b>0,0136</b>	
	Ухиляння від відповіді	<b>0,029</b>	0,0021	
	Дієслівні фрази, які вказує на брак впевненості	<b>0,044</b>	0,006	
	Прикметники (особиста оцінка, а не опис почуттів, речей тощо)	<b>0,062</b>	0,0383	
	Прийменники, які вказують на час, місце розташування тощо	0,0074	<b>0,0279</b>	
	Географічні посилання	0,0018	<b>0,024</b>	
	Стверджувальні фрази	0,0103	<b>0,0286</b>	
	Слова розпоряджень	0,0069	<b>0,0284</b>	
	Дієслова теперішнього часу	0,014	<b>0,036</b>	
	Довгі слова	0,006	<b>0,0091</b>	
	Використання ввічливих слів	<b>0,052</b>	0,019	
	Акцент на словах – "отже", "дуже", "досить" тощо	<b>0,036</b>	0,0109	
	Лайливі та грубі форми слова	0,0089	<b>0,041</b>	
	"Порожні" прикметники	<b>0,0261</b>	0,0045	
	Спеціальна лексика (соціальні, родинні теми дискусій)	<b>0,0306</b>	0,0046	
	Використання умовного способу	<b>0,012</b>	0,0035	
	Уникнення вульгарних і лайливих висловів	<b>0,008</b>	0,0033	
	Підсилювальні слова	<b>0,0224</b>	0,0025	
	Вказівка на розташування об'єктів	0,0099	<b>0,048</b>	
	Займенник першої особи однини (я, мене, моїх тощо)	0,056	<b>0,095</b>	
	Особовий займенник "ми"	<b>0,025</b>	0,009	
	Пестливі слова	<b>0,014</b>	0,0054	
	Спеціальна лексика (спорт, технології, політика, гроші)	0,017	<b>0,084</b>	
Посилання на кількість, величину (згадка суми, кількості)	0,0081	<b>0,035</b>		
Присяги і клятви	0,00063	<b>0,0149</b>		
Згадування минулих подій, які викликають негативні емоції	<b>0,038</b>	0,017		
На рівні речень	Згадування минулих подій відносно теперішнього часу	0,0021	<b>0,039</b>	
	Окличне речення	<b>0,015</b>	0,011	
	Вказівки	0,004	<b>0,046</b>	
	Речення, які висловлюють підтримку	<b>0,0397</b>	0,003	
	Протиставні речення	<b>0,009</b>	0,0023	
	Речення, які слугують для вираження згоди	0,049	<b>0,065</b>	
	Перефразування речень	0,0012	<b>0,0057</b>	
	Непрямі команди і прохання	<b>0,0098</b>	0,0047	
	Підрядні речення обставини	<b>0,098</b>	0,0063	
	Позначення вигуків	0,0046	<b>0,013</b>	
	Безособові речення	0,0097	<b>0,025</b>	
	Підрядні речення, позбавлені змісту	<b>0,0271</b>	0,0056	
	Підрядне речення, яке призначене для уточнення	<b>0,1041</b>	0,0102	
	Заперечні речення	<b>0,0361</b>	0,013	
	Обговорення поточних проблем	0,012	<b>0,044</b>	
	Запитання (вказівки в питальній формі не беремо до уваги)	<b>0,0407</b>	0,0037	
	На рівні речень	Розділові речення	<b>0,0165</b>	0,0056
Прямі цитування		<b>0,005</b>	0,0013	
Еліптичні речення		0,0091	<b>0,0265</b>	
<b>Спільні категорії для обох статей</b>				
На всіх рівнях		Використання жартів	0,041*	0,041*
		Велика кількість слів для написання повідомлення	0,0022*	0,0022*
		Інтимні теми розмови	0,031*	0,031*
	Вираження гніву	0,0188*	0,0188*	

PF – ймовірність появи критерію стилю мови жінки; PM – ймовірність появи критерію стилю мови чоловіка; 0,0n – більша ймовірність появи цієї категорії у повідомленні чоловіка (жінки); 0,0n\* – однакова ймовірність появи цієї категорії у повідомленнях обох статей.

Щоб більшою мірою поставити акцент на написаній інформації ( $P=0,036$ ), особливо на позитивній, жінки використовують у написанні повідомлень велику кількість розділових знаків та декілька знаків оклику чи знаків питань в одному реченні.

Чоловіки описують теперішні події, використовуючи значну кількість дієслів теперішнього часу ( $P=0,036$ ), а жінки згадують минулі події, які викликають негативні емоції ( $P=0,038$ ) та конкретні події, в яких вони брали чи беруть участь тепер.

Щоб висловити свої думки, жінки використовують частіше підрядні речення обставини ( $P=0,098$ ), підрядні речення, які слугують для уточнення ( $P=0,1041$ ), позбавлені смислу ( $P=0,0271$ ), окличні ( $P=0,015$ ) та розділові речення ( $P=0,0165$ ).

Еліптичні ( $P=0,0265$ ) та безособові ( $P=0,025$ ) речення характерніші для чоловічої манери висловлюватись.

Питання ( $P=0,0361$ ) та заперечення ( $P=0,11$ ) – невід’ємна частина комунікації жінок, а ось чоловіків – ствердження ( $P=0,0286$ ), вказівки ( $P=0,046$ ), слів, які висловлюють згоду ( $P=0,065$ ) та розпорядження ( $P=0,0284$ ).

Одними з найяскравіше виражених категорій мови жінки є дієслівні фрази, які вказують на брак упевненості ( $P=0,044$ ), умовний спосіб ( $P=0,042$ ), модальні конструкції ( $P=0,0239$ ), акцент на словах – "отже", "дуже", "досить" та ін. ( $P=0,036$ ), "порожні" прикметники ( $P=0,0261$ ) та прикметники, які вказують на особисту оцінку, а не на опис почуттів, речей тощо ( $P=0,062$ ), протиставні речення ( $P=0,009$ ).

Також чоловіча половина людства переважає з низки аспектів: за довжиною слова ( $P=0,0091$ ), частотою вживання вигуків ( $P=0,013$ ), посиланнями на кількість, величину ( $P=0,035$ ).

Чоловіки обговорюють різні поточні проблеми частіше ( $P=0,044$ ) і більше згадують про відпочинок. Досліджено, що мові чоловіків притаманна наявність згадувань минулих подій відносно теперішнього повідомлення ( $P=0,039$ ).

Для чоловіків мова радше слугує для обміну інформацією про події, об’єкти і процеси ( $P=0,034$ ), вони, обговорюючи теми професій та грошей, використовують спеціальні терміни.

Просторові ( $P=0,136$ ) та географічні ( $P=0,024$ ) посилання, прийменники, які вказують на час, місце розташування тощо ( $P=0,0279$ ), вказівка на розташування об’єктів ( $P=0,048$ ) пріоритетні в повідомленнях, автором яких є чоловік.

Чоловіки більше любляють перефразувати ( $P=0,0057$ ), а ось жінки використовують часто прями цитування ( $P=0,005$ ).

Як показав аналіз, чоловіки у листуванні з іншими чоловіками рідко вживають смайлики, але використовують їх частіше, розмовляючи з жінками, тоді як жінки використовують однакову кількість смайликів в повідомленнях, адресованих як чоловікам, так і жінкам.

Також ми зауважили закономірність використання жарти ( $P=0,0041$ ) в обговореннях: жінки і чоловіки використовують жарти майже однаковою мірою.

Всупереч стереотипам, чоловіки і жінки однаково часто звертаються до інтимних тем розмови ( $P=0,0031$ ), вираження гніву ( $P=0,00188$ ) і використовують в середньому однакову кількість слів для написання повідомлення ( $P=0,0022$ ).

Потрібно зауважити, що на результати досліджень істотно впливають як контекст повідомлень, так і теми дискусій. У нашому дослідженні ми цей факт взяли до уваги і зробили різнопланову гендерну вибірку повідомлень користувачів з всіх тематичних розділів "Форуму Рідного Міста". У розділах "Форуму Рідного Міста" ми однаковою мірою розглянули дискусії, які виникають у зв’язку з різноманітними інтересами та захопленнями як чоловіків, так і жінок.

#### 4.4 Соціальні діалектизми

Соціальні діалектизми (соціолекти) – слова, притаманні певній соціальній групі. Здебільшого соціолекти різняться лише лексикою (а саме: іменниками, прикметниками, дієсловами і прислівниками).

Дослідження лексики соціолекту [23] дають можливість виокремити всередині соціуму групи людей за певними ознаками – професією, родом діяльності, інтересами, віком тощо. Для визначення професії, віку і територіальної приналежності ми проаналізували лексику обмеженого соціального вжитку.

*Професіоналізми* — слова, які вживаються для позначення спеціальних понять у сфері тієї чи іншої професії. Наприклад:

- у мові моряків (*кок, камбуз, кубрик тощо*);
- у водіїв (*баранка, двірники тощо*);
- у футболі (*свічка, дев'ятка, аут тощо*);
- у галузі шахтарської справи (*штрек, терикон, видавати на гора тощо*);
- у сфері комп'ютерної технології (*курсор, дискет, глосарій, вікно редактора, вінчестер, командний рядок, файл тощо*);
- у мовленні мисливців, рибалок (*поплавок, підсака, курок, гачок, трофей тощо*);
- працівників банківсько-фінансової, торговельної та подібних сфер (*зняті каси, підбити, прикинути баланс тощо*);
- у музикантів (*фанера; розкрутити, сольник*);
- у медиків (*скальпель, операційна, реанімація тощо*) тощо.

Професійно диференційовані соціолекти дають змогу зробити припущення про коло інтересів і професії автора повідомлення.

#### 4.5. Безеквівалентна лексика

Наступним критерієм ідентифікації користувача є національність. Дослідження відмінностей мов, зумовлених своєрідністю культури, зводяться до дослідження відмінностей у лексиці мов. Саме результати дослідження безеквівалентної лексики [23], до якої належать слова, що позначають специфічні явища культури і не мають однослівного перекладу на іншу мову, дадуть можливість визначити національність користувача.

*Екзотизми* – безеквівалентні слова у лексиці людини, які при перекладі передаються описово або запозичуються.

Наведемо приклади екзотизмів – слів, які пов'язані з матеріальною і духовною культурою певної країни: англійська безеквівалентна лексика – *стерлінг, біг-бен, шилінг, крикет*; японська – *сакура, гейша, ікебана, саке, кімоно*; італійська – *конклав, сентимо, спагеті, тарантела*; іспанська, – *песо, конкістадор, тореадор, корида, болеро*; американська – *ковбой, рейнджерси, авеню, барбі*; російська – *сарафан, щі, боярин, балалайка, ямицьк*; середньоазіатська – *кишлак, арик, дехканин, кобуз, домбра*, до української безеквівалентної лексики належать такі слова, як *чумацьк, гривня, ушник, галушки, вареники, борщ, бандура, кобзар, коломийка, гопак, вечорниці, тризуб*.

*Територіальні діалектизми* – група слів, поширення яких обмежується певною територією. Наприклад:

- картопля – *бараболя, крумплі, бульба*;
- бур'ян – *хоптя*; відро – *путня*;
- димар – *верх*;
- гарна погода – *година* тощо.

Визначення територіальних діалектів, місцевих реалій і понять (не відомих або не використовуваних поза межами певного говору чи групи споріднених говорів) у повідомленнях користувача дає змогу визначити область і місто, де він проживає.

Наприклад, якщо в мові листування було використано реалії – *легінь, ватра, кептар, колиба, трембіта, газда*, то велика ймовірність того, що користувач проживає чи виріс у Карпатах (Івано-Франківській чи Закарпатській області).

#### 4.6. Визначення ймовірності авторів повідомлень у Інтернет-спілкуванні.

Для визначення ймовірності ідентифікації авторів повідомлень у інтернет-спілкуванні нам спочатку потрібно встановити найімовірнішу кількість появ досліджуваних словоформ (слів-ознак певного стилю мовлення) в текстовому наповненні форуму.

Оскільки основою всіх індуктивних досліджень у лінгвістиці є спостереження за лінгвістичними об'єктами, яке здійснюється через кількісне вимірювання, ми присвоїмо досліджуваному слову (або словоформі) змінну  $S$ , яка з'являється у вигляді слововживань у текстовому наповненні форуму  $F$  разів.

Найімовірнішу кількість появ досліджуваних словоформ у дискусіях користувачів знайдемо за класичним означенням ймовірності [24]:

$$P(S) = \frac{F}{N}, \quad (1)$$

де  $S$  – досліджувана словоформа;  $F$  – кількість слововживань досліджуваних словоформ в текстовому наповненні форуму;  $N$  – загальна кількість слововживань словоформи в текстовому наповненні форуму,  $0 \leq P(S) \leq 1$ .

Модальне значення визначимо за формулою [24]:

$$Np + q \leq x_0 \leq Np + p, \quad (2)$$

де  $q = p - 1$ ,  $x_0$  – модальне значення,  $p \neq 0$  і  $p \neq 1$ ,  $N$  – кількість слововживань словоформи в текстовому наповненні форуму;  $P(S) = p$ ,  $p$  – ймовірність появи словоформи в кожному окремому повідомленні;  $q$  – ймовірність не появи словоформи в кожному окремому повідомленні.

Таким методом можна обчислити ймовірність появи словоформ, які належать до сформованих нами категорій мови у реченні/повідомленні різної довжини.

Отже, розрахуємо ймовірність появи словоформ стилю мови жінки у повідомленні користувача “Львів. Форум Рідного Міста” пані Сумної (<http://misto.ridne.net/viewthread.php?tid=2453>).

У табл. 3 і 4 подано результати аналізу цього повідомлення на рівні речення і слова.

Таблиця 3

#### Аналіз повідомлення на рівні речення

<i>Речення, які використовують у своїх повідомленнях жінки</i>	<i>Приклад</i>	<i>N</i>	<i>P-повід. п. Сумної</i>	<i>PF-контр. значення</i>
Речення, які вказують на брак впевненості	<i>Я, напевно, ще не все згадала.</i>	2	0,05	0,044
Підрядні речення, позбавлені змісту	<i>На мою думку, зрозуміти іншого взагалі неможливо.</i>	1	0,026	0,0271
Підрядне речення, яке служить для уточнення	<i>Розкіш бути собою: не намагались сподобатися, не прикидатися кішечкою.</i>	5	0,128	0,1041
Прямі цитування	<i>Моє кредо – “Промовляння самотній” Ю. Андруховича.</i>	2	0,051	0,005
Підрядні речення обставини	<i>Усвідомлюю, що мене чекає те ж, що і біблійне дерево, яке не приносить плоду.</i>	5	0,128	0,098
Заперечне речення, твердження про відсутність чогось	<i>Жодних подруг, друзів, розваг тощо</i>	6	0,154	0,11
Запитання	<i>Може, мені справді добре?</i>	4	0,103	0,0407
Розділові речення	<i>Хоча творчій людині треба і його спізнати, чи не так?</i>	1	0,026	0,0165
Використання умовного способу	<i>Хотілось би знати, як себе почувають інші самотні серед несамотніх.</i>	2	0,051	0,042
Вираження подяки	<i>Дякую за такі доброзичливі і щирі дописи.</i>	2	0,051	0,063

## Аналіз повідомлення на рівні частин мови, конкретних слів та фраз

Слова, які використовують у своїх повідомленнях жінки	Приклад	N	P- повід. п. Сумної	PF-контр. значення
Невпевненість	<i>чи, намагатись, щось</i>	8	0,02	0,045
Заперечення	<i>не, ніяких, ніколи</i>	24	0,057	0,11
Емоції та почуття	<i>свобода, роздратованість, заздрість</i>	28	0,06	0,036
Ухиляння від відповіді	<i>мабуть, здається</i>	4	0,02	0,029
Прикметники (особиста оцінка)	<i>бозна-яка, корисних, найгірша</i>	27	0,06	0,062
Акцент	<i>отже, отож</i>	4	0,02	0,036
Згадування минулих подій відносно теперішнього часу	<i>оминула, зробила</i>	10	0,027	0,039
Спеціальна лексика (соц., родинні теми)	<i>розвага, друзі, лахи, прикраси</i>	12	0,029	0,0306
Уточнення	<i>що, щоб, коли, яке</i>	13	0,028	0,1041
Питання	<i>чому, чого</i>	4	0,019	0,0407
Умовний спосіб	<i>б, би</i>	2	0,01	0,012
Модальність	<i>можу, потрібно</i>	3	0,01	0,0239
Підсилювальні слова	<i>особливо, справді, навіть</i>	12	0,028	0,0224
Слова подяки	<i>дякую</i>	2	0,01	0,063
Пестливі слова	<i>кішечка</i>	1	0,007	0,014
Негативні слова минулого часу	<i>конфліктував, сумувала</i>	11	0,026	0,038

На підставі отриманих результатів аналізу повідомлення п. Сумної на рівні речення і слова можна констатувати, що є невеликі розбіжності між ймовірностями появи сформованих нами категорій, які характерні для мовлення жінки, у повідомленні п. Сумної, та контрольними значеннями. Наведені результати свідчать, що контекст і тема повідомлення істотно впливають на результат аналізу.

### Висновки

Щоденно зростає кількість користувачів соціальних мереж, форумів, чатів тощо. Це породжує проблему перевірки правдивості наданих даних користувачів цих веб-спільнот. Розроблення та дослідження комп'ютерно-лінгвістичних методів перевірки достовірності наданої інформації користувачем, які полягають у визначенні особливостей стилів написання повідомлень користувачів веб-спільнот, дає змогу вирішити цю проблему.

Результати нашого дослідження відмінностей стилів мови у інтернет-комунікації можуть істотно допомогти в створенні ефективних методів перевірки достовірності особистої інформації, яку надав користувач. Ми проаналізували інформаційне наповнення чотирьох веб-форумів з різною тематикою і цільовими групами дописувачів (а саме: "Львів. Форум Рідного Міста", "Тернопільський Форум", "Родинний форум "Малеча", "ROCK.LVIV.UA: Форум"). Дослідження показало, що ми можемо багато довідатися про людей: характер, звички, емоції і мотиви, підрахувавши відмінності стилів написання ними повідомлень у дискусії веб-спільнот та класифікації слів, речень, які вони використовують у цих комунікаціях. Проаналізувавши останні дослідження, ми розробили основу для методики комп'ютерно-лінгвістичної перевірки достовірності наданої інформації користувачів мережі у своїх аккаунтах та визначення ймовірності вживання характерних для певного стилю мовлення категорій у інтернет-спілкуванні. В подальших дослідженнях потрібно здійснити поглиблений вибір цих категорій та класифікувати їх за спільними характеристиками. Особливу увагу ми приділили віковим та гендерним відмінностям вживання мови під час написання повідомлень в онлайн-комунікації. Цей лінгвістичний матеріал є доброю основою для розроблення спеціалізованого програмного забезпечення, яке за допомогою комп'ютерно-лінгвістичних методів перевірки достовірності наданої інформації користувачем дає змогу визначити особистість автора опублікованих у мережі повідомлень.

1. Argamon S. *Gender, Genre, and Writing Style in Formal Written Texts* / Argamon S., Koppel M., Fine J., Shimon A. – Text, volume 23(3), 2003. – P. 321–346. 2. Constantin C. *Impression Formation Effects in Moderated Chatrooms: An Experimental Study of Gender Differences* / Constantin C., Kalyanaraman S., Stavrositu C., Wagoner. – LA: New Orleans, 2005. – [Електронний ресурс] – Режим доступу – <http://www.psu.edu/dept/medialab/research/NCA.htm>. 3. Corney M.W. *Analysing E-mail Text Authorship for Forensic Purposes*, Master Thesis / Corney M.W. – Queensland, 2003. 4. Crawford M. *Talking Difference: On Gender and Language* / Crawford M. – London: Sage Publication, 1995. 5. Crystal D. *Language and the Internet* / Crystal D. – Cambridge: Cambridge University Press, 2001. 6. Dubois E. *The Question of Tag Questions in Women's Speech: They Really Don't Use More of Them.* / Dubois E., Crouch I. – Language in Society 4, 1975. 7. Henley N. *Gender, power and miscommunication. Miscommunication and Problematic Talk* / Henley N., Kramarae Ch. – Newbury Park: Sage, 1991. – [Електронний ресурс] – Режим доступу – [http://www.ji.lviv.ua/n35texts/masenko-usni\\_formy.htm](http://www.ji.lviv.ua/n35texts/masenko-usni_formy.htm). 8. Herring S.C. *Politeness in computer culture: Why women thank and men flame* // Cultural Performances / Herring S.C – Proceedings of the Third Berkeley Women and Language Conference, 1994. – [Електронний ресурс] – Режим доступу: Berkeley Women and Language Group: <http://ella.slis.indiana.edu/~herring/politeness.1994.pdf>. 9. Lakoff R. *Language and Woman's Place.* / Lakoff R. – New York: Harper & Row, 1975. 10. Mulac A. *Female and male managers and professionals criticism giving* / A. Mulac, D.R. Seibold, J.L. Farris // Differences in language use and effects Journal of Language and Social Psychology, Vol.1, 2000. – P.389–415. 11. Mulac A. *Linguistic contributors to the gender-linked language effect* / A. Mulac, T.L. Lundell // Journal of Language and Social Psychology, 5(2), 1986. – P.91–101. 12. Mulac A. *Male/female language differences and attributional consequences in children's television* / A. Mulac, J.J. Bradac, S.K. Mann // Human Communication Research, 11(4), 1985. – P.481–506. 13. Reid E. *Electropolis: Communication and community on the Internet Relay Chat.* BA Honors Thesis // Department of History, University of Melbourne / E. Reid. – Australia: Melbourne, 1991. 14. Rodino M. *Breaking out of Binaries: Reconceptualizing Gender and its Relationship to Language in Computer-Mediated Communication* // Journal of CMC 3 (3) / Rodino M. – 1997. 15. Shuttleworth J. *Living Language* / J. Shuttleworth, G. Keith. – London: Hodder Murray, 2000. 16. Tannen D. *You just don't understand: Women and men in conversation* / D. Tannen. – New York: Ballantine Books, 1991. 17. Thomson R. *Predicting Gender from Electronic Discourse* / R. Thomson, T. Murachver // British Journal of Social Psychology. – vol. 40. – № 2. – 2001. 18. Walther J.B. *The impacts of emoticons on message interpretation in computer-mediated communication* / J.B. Walther, K.P. D'Addario // Social Science Computer Review 19, 2001. – P.323–345. 19. Wikipedia, the free encyclopedia. – Режим доступу до енциклопедії: <http://en.wikipedia.org/wiki/>. 20. Witmer D.F. *Smile When You Say That: Graphic Accents as Gender Markers in Computer-Mediated Communication* / D. Witmer, S. Katzman. – CA: Menlo Park, 1997. 21. Witmer D.F. *On-line Smiles: Does Gender Make a Difference in the Use of Graphic Accents* / D.F. Witmer, S.L. Katzman // Journal of Computer-Mediated Communication. – vol. 2, № 4. – 1997. 22. Дзюбишина-Мельник Н.Я. *Сучасний жаргон й сучасне розмовне мовлення* // Наукові записки Національного університету «Києво-Могилянська Академія». – Т. 20. Філологічні науки. / Н.Я. Дзюбишина-Мельник. – К., 2002. 23. *Курс лекцій з лексикології, фразеології та лексикографії.* / Склад лексики української літературної мови з погляду її походження. Режим доступу – <http://www.tspu.edu.ua/subjects/814111/leksykology/lec.htm>. 24. Пиотровский Р.Г. *Математическая лингвистика* / Р.Г. Пиотровский, К.Б. Бектаев, А.А. Пиотровская. – М.: Высшая школа, 1977.