

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Київський національний університет імені Тараса Шевченка
Навчально-науковий інститут філології

Кафедра української мови та прикладної лінгвістики

Корпус суржикових текстів

Кваліфікаційна робота

освітнього ступеня «бакалавр»
за спеціальністю 035 «Філологія»,
спеціалізацією 035.10 «Прикладна
лінгвістика»,
галузі знань 03 «гуманітарні науки»
ОПП «Прикладна (комп'ютерна)
лінгвістика та англійська мова»
студентки IV курсу

Марини МАЖАРСЬКОЇ

Науковий керівник:

к.філол.н., доц. Оксана ЗУБАНЬ

Рецензент:

д.філол.н., проф. Наталія ДАРЧУК

«Допущено до захисту»

Протокол № 11 засідання кафедри
української мови та прикладної лінгвістики

ННІФ від 01.06.2023

Завідувач кафедри _____ **Сергій Різник**

КИЇВ – 2023

ЗМІСТ

ВСТУП	3
РОЗДІЛ I. ТЕОРЕТИЧНІ ЗАСАДИ ТРАКТУВАННЯ ЯВИЩА СУРЖИКУ У ЛІНГВІСТИЦІ	6
1.1. Тракткування поняття «суржик» у лінгвістиці	6
1.2. Причини та період закріплення суржиків у мовленні українців.....	11
1.3. Суржик та білінгвізм в аспекті інтерференції.....	16
1.4. Використання суржиків в українськомовному соціумі.....	18
1.5. Лінгвістичні ознаки суржиків	22
1.5.1. Типізація суржикового мовлення	22
1.5.2. Фонетичні ознаки суржиків	23
1.5.3. Лексичні особливості суржиків	26
1.5.3. Синтаксичні особливості суржиків	27
Висновок до розділу I.....	28
РОЗДІЛ II. КОРПУС СУРЖИКОВИХ ТЕКСТІВ	30
2.1. Етапи укладання корпусу суржикових текстів	33
2.1.1. Збір аудіозаписів	33
2.1.2. Розпізнавання усного мовлення.....	34
2.1.3. Обробка записів у програмі Praat та створення TextGrid	35
2.1.4. Текстове представлення аудіозаписів	38
2.1.5. Формування реєстру суржикізмів.....	39
2.1.6. Лінгвістичне опрацювання суржикізмів	40
2.1.7. Укладання бд корпусу суржиків та її структура	41
2.2. Статистична параметризація суржикового мовлення за матеріалами корпусу.....	46
Висновок до розділу II	51
ВИСНОВКИ	53
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	56
ДОДАТКИ	61

ВСТУП

Сьогодні існує багато наукових праць, які розглядають таке мовне явище, як суржик. Зокрема, праці Масенко Л. [21], Радчука В. [27], Селіванової О. [26], Біланюк Л. [41], Тараненко К. [32], Томіленко Л. [34] та інших. Але, незважаючи на це, досі не існує усталеного трактування того, що таке суржик. Загалом, це явище можна окреслити як суміш української та російської мов, але деякі лінгвісти відносять суржик до окремих різновидів, наприклад, діалектів, піджинів, «недоперекладу» тощо. Відсутність чіткого визначення впливає також і на висновки щодо інших питань: причини, передумови, риси цього явища і, насамперед, питання того, які тексти вважати суржиковими. Звідси виникає проблема збирання текстового матеріалу для його систематизації, а відсутність систематизованого матеріалу породжує неможливість детального вивчення суржику як мовного явища. До того ж, суржик побутує переважно в розмовному мовленні, а тому може існувати декілька форм одного і того ж слова.

Інший бік цього питання – у якій формі доречно збирати текстовий матеріал для таких досліджень? Найзручнішим способом представлення певної мови чи мовних варіантів є корпус. Зазвичай, корпус у лінгвістиці розглядають як вибірку комп'ютерних репрезентативних текстів, що підібрані, упорядковані та розмічені відносно різних критеріїв. Корпусами зручно користуватися завдяки комп'ютерному представленню інформації – можна фільтрувати тексти, а також здійснювати пошук конкретних слів або їх особливостей. Вважаємо, що саме корпус у контексті дослідження суржикового мовлення є найкращим способом його систематизації. У корпусах міститься різна метатекстова та лінгвістична інформація про окремі його структурні елементи. Це дає змогу не просто зібрати словник словоформ чи лексем суржикізмів, а також подати й іншу інформацію – текстовий контекст вживання, інформацію про ситуативне вживання суржику, нормативний відповідник та багато іншої. Уся вона є суттєвою для вивчення суржику, оскільки, як згадувалося раніше, досі не існує репрезентативної збірки для відображення особливостей цього мовного явища. *Актуальність*

проведеного дослідження пояснюється, по-перше, необхідністю лінгвістичного обґрунтування і всебічного вивчення явища суржику, по-друге, відсутністю текстової бази для дослідження цього явища, зокрема, створеного та розміченого корпусу суржикових текстів.

Метою роботи є укладання корпусу суржикових текстів та його аналіз.

Реалізація поставленої мети передбачає вирішення низки **завдань**:

- опрацювати лінгвістичну літературу за визначеною темою, обґрунтувати поняття суржику та розробити типологію суржикізмів;
- відібрати аудіозаписи для роботи, створити текстґриди до них із необхідними позначками (зокрема, для вказівки на суржикізм);
- подати текстове представлення усіх аудіозаписів;
- створити реєстр суржикізмів;
- здійснити лінгвістичне опрацювання отриманих суржикових словоформ;
- створити базу даних суржикового корпусу;
- провести статистичну параметризацію за матеріалами укладеного корпусу.

Об'єктом дослідження є тексти (аудіозаписи та їх письмові варіанти) суржикового мовлення, а **предметом** – лінгвістичні ознаки (фонетичні, лексичні, граматичні) суржикової лексики.

Матеріалом дослідження виступають аудіозаписи із особистих розмов, що містять суржикові слова. Для дослідження відібрано 135 аудіозаписів формату .wav різної тривалості — від 15 секунд до 5 хвилин. Загальна тривалість усіх оброблених аудіозаписів становить годину сорок хвилин та сорок секунд, тобто, приблизно 6041 секунду усного мовлення. Усі аудіозаписи містяться у [папці](#), їх назвами є цифри від 1 до 135, що відображає порядок їх опрацювання.

Методи дослідження: експериментально-фонетичні методи, статистичні методи, методика графічного представлення статистичних даних.

Новизна та практична значущість дослідження полягають у створенні корпусу суржикових текстів, який надалі можна використовувати для лінгвістичних досліджень. У корпусі містяться два словники – лексем та словоформ. Словник лексем пізніше можна автоматично розширити, вивівши усі словоформи лексеми. Словник словоформ підходить для дослідження частоти та вживаності певних форм слова. Анотовані записи та транскрипції словоформ можна використати для навчання систем розпізнавання мовлення, тому що суржик погано розпізнається та знижує відсоток успішності розпізнавання. Нормативний варіант, підібраний до словоформи, можна використати для автоматичного виправлення помилок у текстах.

Структура кваліфікаційної роботи. Робота складається із вступу, двох розділів основної частини, висновків до кожного з розділів, загальних висновків, списку використаної літератури та списку додатків.

РОЗДІЛ І. ТЕОРЕТИЧНІ ЗАСАДИ ТРАКТУВАННЯ ЯВИЩА СУРЖИКУ У ЛІНГВІСТИЦІ

1.1. Тракткування поняття «суржик» у лінгвістиці

Українська мова є державною мовою в Україні, її статус та права закріплені Конституцією. Але серед населення також поширені інші мови – зокрема, російська. Також частково поширені й інші мови національних меншин: білоруська, румунська, угорська, польська, кримськотатарська та інші. Серед цього мовного різноманіття має місце і суржик. Суржик – мовне явище, що поширилося у мовленні українців під впливом російської мови на українську. Упродовж декількох століть російська мова була домінантною на території української держави, особливо на Сході та Півдні України, а також у центральних регіонах. Загалом, суржик менш поширений на Заході – це пояснюється меншим історичним впливом Росії на цю частину України.

Питання походження, розвитку та визначення самого поняття «суржик» є неоднозначним. Найпоширеніше тлумачення для слова «суржик» закріплене у Словнику української мови у 11-ти томах: «1. Суміш зерна пшениці й жита, жита й ячменю, ячменю й вівса і т. ін.; борошно з такої суміші; 2. *перен., розм.* Елементи двох або кількох мов, об'єднані штучно, без додержання норм літературної мови; нечиста мова» [29, Том 9, 1978., с.854]. Таке визначення є популярним як серед лінгвістів, так і серед звичайних людей, зокрема, його подають у багатьох статтях різних засобів масової інформації. Воно і відображає в загальному те, що таке суржик.

Суржик є серйозною проблемою, яка шкодить українській літературній мові. Ненормативні форми запам'ятовуються й поширюються серед населення та закріплюються у мовленні. Саме тому, багато науковців та лінгвістів звертають свою увагу на вивчення цього явища. У результаті – варіативність поглядів на те, що ж таке суржик.

Деякі науковці вважають, що суржик можна віднести до явищ, що виникли завдяки взаємодії мов. До них належать піджини, креольські мови, койне, змішані мови тощо. Ці мовні явища функціонують як засіб спілкування між населенням територій, що комунікують різними мовами або діалектами, їх основна задача – полегшення комунікації.

Піджин – проста мова, із невеликим словниковим складом (лексика спрощена та викривлена) та простою граматиною. Її основа формується на граматиці мов народів, які комунікують між собою. Вона не є рідною для жодного з народів і використовується як друга, «іноземна» мова. [8, с. 151]. Піджини не існують тривалий час. Коли зникає причина їх появи та потреба у використанні, вони поступово занепадають. Однак, інколи піджин стає основою для створення креольських мов. Креольська мова тоді функціонує як самодостатня. Найбільше креольських мов сформувалося на островах на основі англійських, португальських, іспанських та французьких піджинів.

Думки щодо віднесення суржику до піджинів або креольських мов різняться. Наприклад, Селіванова О. зазначає: «На нашу думку, суржик є формою міксації двох споріднених мов, подібним до одного з різновидів піджину і результатом окресленого вище ціннісного конфлікту. Однак, як відомо, піджин є, насамперед, соціальним субстратом, мовою спілкування певної соціальної групи, суржик же характеризує мовлення жителів села, що переїхали до міста в першому поколінні. Якщо зважати на те, що ці люди є певною соціальною стратою, то суржик є типом піджину на етапі створення, хоч і не набуває статусу засобу міжетнічного спілкування» [26, с. 7]. Загалом, позиція дослідниці не є однозначною – вона вважає, що суржик частково можна включити до піджинів, але водночас з цим спирається на відсутність конкретних змін у граматиці мов при творенні суржику. Це дійсно так, адже найперша умова функціонування піджину – побудова простої граматичної системи, що не буде важкою для засвоєння обома народами, які повинні користуватися цим мовним засобом. Але

для суржиків окремої простої граматичної системи не існує. Більшість російських слів, які потрапляють у суржик, зазнають фонетичних та морфологічних викривлень у відповідності до норм української мови. Крім цього, виникає суперечливість, якщо розглядати процес закріплення піджину та його перетворення на креольську мову. Зазвичай, місцеве населення має дуже слабкі мови, вони не є унормованими у граматиці, а мова іншого народу навпаки є домінантною. Суржик, у свою чергу, формується на основі двох самостійних мов, статус та правила яких є закріпленими. Українська та російська мови мають свої мовні стандарти, які постійно змінюються та підлаштовуються під потреби мовців. Підтримати думку дослідниці також можливо, адже суржик як засіб комунікації використовують із ціллю «бути зрозумілим», що, в цілому, відповідає концепту появи піджину як мовного засобу. Але, незважаючи на це, все одно важко віднести суржик до піджинів та креольських мов, тому що основні умови його функціонування суперечать ідеї витворення цих мовних явищ.

Деякі дослідники відносять суржик до просторіччя. За даними Словника української мови в 11 томах, просторіччя – це «слово, граматична форма, зворот, що не є нормою літературної мови, але вживається в побутовому мовленні, а в художніх творах використовується як стилістичний засіб для надання мові зневажливого, іронічного, жартівливого, грубуватого та ін. відтінку» [29, Том 8, 1977., с. 301]. Думки, що суржик належить до просторіччя, притримується Труб В.: «Цей вид [суржикова форма українського просторіччя] суржиків природно кваліфікувати як форму просторіччя чи просто-розмовної мови, яка утворюється в ситуації нерівноправної (диглосної) двомовності» [33, с. 143]. У цілому, дійсно можливо розглянути суржик як елемент просторіччя, тому що він функціонує у побутовому мовленні українців як засіб їх спілкування між собою. Але просторіччя – явище лише однієї мови, у межах її власних фонетичної, лексичної та граматичної систем. У суржиків ці системи перетинаються для двох окремих

мов, тому відбуваються різного роду викривлення, зокрема, і для слів, що дійсно є просторіччям, наприклад, «дрихнути» (просторічна форма для «спати») – «дрихнуть» (суржиковий елемент).

Суржик як недопереклад розглядає Радчук В.: «Справді, суржик інтересний тим, що по ньому вгадуєш, з якої мови робився переклад. А краще б сказати, з якої мови переклад недороблено. Межа між суржигом і нормою, надто між варваризмами і узаконеними запозиченнями, хистка й непостійна. І за відсутності належної лексикографії та освіти кожен проводить її на свій розсуд. А проте навіть нечіткої межі часом досить, щоб углядіти вторинність тексту і безпомилково визначити мову, з якої здійснено недопереклад.» [27, с. 12]. Варто зазначити, що стаття дослідника – це критика іншої статті. У ній він також згадує «українсько-польський», «українсько-англійський» та навіть «українсько-скіфський» суржик, а також не вбачає великої різниці між функціонування англіцизмів і суржикізмів у мовленні українців. Продовжуючи думку про різні види суржику, згадаємо й визначення укладачів енциклопедії «Українська мова». Вони вважають, що термін суржик можна використовувати для характеристики різних змішаних типів мов, наприклад, українсько-польського, українсько-словацького, українсько-англійського та українсько-російського, але зазначають, що найчастіше суржик – це українське просторіччя із домішкою русизмів, що використовується як розмовна мова [36, с.665].

Закордонні лінгвісти у своїх працях також приходять до різних висновків щодо поняття «суржик». Катерина Кент відносить суржик до категорії «українсько-російських мовних перемикань» (тут і далі переклад наш. – М. М.): «Я вважаю, що це випадок мовного російського запозичення в українську, хоча й спланованого. Я б класифікувала «суржик» міських білінгвів як випадок українсько-російського мовного перемикання. Як впливає з назви цього різновиду «суржику», його носії володіють як українською, так і російською мовами.» [45, с. 29]. Мовне перемикання – процес перескакування від однієї мови

до іншої в межах розмови. Зазвичай, це вимагає гарного знання обох мов і усвідомлення того, коли яку мову використовувати краще. Класифікація суржику як мовного перемикування не є цілком доречною, адже, знову ж таки, у системі суржикового мовлення відбуваються суттєві спотворення слів на різних мовних рівнях. Хентшел Г. неодноразово згадує суржик як «варіант» або «розмовну мову» (тут і далі переклад наш. – М. М.): «В українському варіанті російської мови можуть бути присутні, звичайно, фонетичні особливості або специфічні слова українського походження. Такий місцевий колір є нормальним для розмовних варіантів європейських стандартних мов. Проте значна частина українців зберегла суржик як неофіційну народну мову, принаймні для приватної сфери» [43, с. 3].

Існують й інші бачення цього терміна: «мовне перемикування» (Саплін Ю.), «субкод» (Масенко Л.), «напівмовність», «ноутворена мовна система» (Брацкі А.), «мішана мова», «мовний покруч». Дехто вважає суржик своєрідним діалектом, а його носіїв – діалектоносіями [34, с. 71 - 73]. Зокрема, Дель Гаудіо С. вважає, що існують різні варіанти суржику, тому що існують різні діалектні варіанти українських слів. Дослідник позначає суржик як наддіалектне явище, тому що він не використовується в залежності від регіону, а є характерним для всієї території України. В одному випадку суржикізм може бути дійсно суржикізмом, тобто викривленим російським словом, а у іншому випадку – це саме слово вже не буде суржикізмом, а лише діалектом певної місцевості [28, с. 140].

Загалом, можна зробити висновок, що на сьогодні не існує єдиного визначення того, що таке суржик. Однак, серед усіх дослідників, науковців та лінгвістів поширена думка про те, що суржик – це суміш мов (зазвичай російської та української), яка не є унормованою та достатньо дослідженою для віднесення до якоїсь однієї категорії.

1.2. Причини та період закріплення суржик у мовленні українців

Спочатку суржик був проявом переважного усного мовлення, особливо на територіях, на які Росія мала значний вплив. Російська мова використовувалась як офіційна у різних сферах, а отже, люди були змушені розуміти її та розмовляти нею. Сучасний суржик поширився й на інші сфери. Приклади його використання можна зустріти на телебаченні та в пресі («Коли я заходжу на Оболоні в супермаркет і чую пісню *«Мой адрес – не дом і не улица, мой адрес – Советский Союз»*, я потрапляю в *Советский Союз* або в *російську імперію.*» (Ірина Халупа, Радіо «Свобода», аудіозапис програми «30 хвилин у різних вимірах: суржик» від 04.06.2005)), а також у різних літературних («*Не, ні фіга. Треба натиснути правою кнопкою на слово «Лінукс» і вибрати «добавить в словарь»*») [37] та пісенних текстах («*Суботній вечір, в повітрі запах духів / На твої плечі я свою куртку одів / І потім довго вона тримала твій смак / Я до сьогодні не міг забути ніяк / Наш самий перший медляк!*» (Кузьма Скрябін «Наш перший медляк»)).

Які ж причини виникнення такого явища? Белей Л. виділив такі передумови для формування нових «кодів» (на думку автора, на позначення суржика «доцільно вживати нейтральний термін «код», що не є ані мовою, ані – соціолектом» [3, с.6]):

1. тісний зв'язок між субстратом та суперстратом (колонізація території України Росією впродовж багатьох століть; насильницьке переселення українців та заселення звільнених територій росіянами);
2. суперстрат є основною мовою спілкування для більшості населення (російською мовою спілкувалась значна частина Російської імперії та СРСР);
3. носії субстрату мають лише одну рідну мову та не відрізняють інші мови від неї;

4. суперстрат та субстрат контактують протягом тривалого періоду часу (російська мова почала впливати на українську ще з кінця XVII ст.) [3, с. 6].

Томіленко Л., спираючись на історичні факти, теж виокремлює ряд причин появи суржику:

1. захоплення Російською імперією центральних та південно-східних територій України;
2. уведення російської мови в основні сфери життя та неодноразова заборона використання та поширення української мови;
3. переселення та депортація народів;
4. територіальна та мовна близькість тощо [34, с.69]

За словами Германа В., суржик виник як наслідок довготривалої політики лінгвоциду, а також через те, що російська мова вважалася пріоритетною та «вищою» мовою [9, с. 41]. Серед інших причин також виділяють заниження статусу української мови, сприйняття її як «низької», двомовність сім'ї. Білінгвізм на фоні спорідненості мов також є однією з можливих умов формування цього явища.

Загалом, дослідники виділяють доволі схожі причини виникнення суржику, що є очевидним, адже саме вплив Російської імперії та російської мови на українців є першопричиною міксації мов між собою.

Масенко Л. поділяє історію виникнення українсько-російського змішаного мовлення на такі періоди:

1. Кінець XVII – початок XVIII століть. У цей період російська мова була основою для ключових сфер життя, зокрема адміністративної. Це призвело до того, що серед чиновництва сформувалося усне мовлення, яке поєднало у собі російський канцелярит та елементи літературної української мови

[21, с. 13]. Приклади його використання можна знайти у різних літературних творах того часу. До прикладу, цитата із поеми «Сон (комедія)» Тараса Шевченка: «Де ти здесь узялся?» / «З України». — «Так же ти / Й говорить не вміш / По-здешньому?» — «Ба ні, — кажу, — / Говорить умію, / Та не хочу». — «Экой чудак! / Я всі входи знаю, / Я тут служу, коли хочеш, / В дворец попытаюсь / Ввести тебе. Только, знаєш, / Мы, брат, просвищенны, — / Не поскупись полтинкою...» [38].

Суржик також поширився і серед військових, які йшли на службу. Проблема полягала в двомовності – на службі доводилося спілкуватися російською мовою, а в колі сім'ї та друзів було прийнято говорити українською. Через це відбувалось змішування цих мов. Таким чином, свій початок суржикове мовлення, загалом, взяло не від селян, а від управлінського та військового прошарку населення.

2. Кінець XIX століття – початок XX століття. Через пришвидшені темпи індустріалізації та урбанізації багато людей змінювали своє місце проживання, обираючи місто. В українських містах у цей період проживала лише третина українців, частина росіян та євреїв, а інші жителі були переважно зрусифікованими. [21, с. 18] Наслідком цього стало розширення меж використання суржику, оскільки тепер його використовували не тільки чиновники та солдатство, а й селяни, що переїхали до міст і працювали у сфері промисловості.
3. Період політики коренізації та українізації. Коренізація – політика більшовицького уряду, направлена на розширення сфер вживання національних мов республік. В Україні вона отримала назву «українізація». Українізація відзначилася поширенням української мови в управлінській діяльності. У період з 1923 до 1927 рр. для політичних діячів проводили курси української мови та змушували їх вести документацію

саме нею. [13, с.61]. Але з часом ця політика була припинена, тому що суттєво впливала на державну ідентичність українців.

Суттєвою причиною закріплення суржику вже у письмовому, а не тільки у розмовному вигляді стало просування гасел «братніх народів» та «спорідненості мов». Крім заборон використання української мови, уряд також почав процес зміни структури української мови, тобто штучного поєднання лексики та граматики української та російської мов. Дослідження цього питання є у праці «Радянська мовна політика в Україні. Переклади творів Леніна українською 1930-х та 1950-х рр.» Гофенедера Ф. Дослідник подає приклади лексичних замінів, що проводив радянський уряд. «Як показав аналіз, у виданні 1952 року є чимало форм, що невдовзі склали основу суржику. Хоча ця гіпотеза й потребує додаткових доказів, та припускаємо, що масове поширення суржику відбулося саме через книжки (в тому числі твори Леніна) та інші видання, які з'явилися в проміжку між 1932-м – роком завершення політики українізації – і відлигою кінця 1950-х. Як наслідок, суржик, сферу існування якого переважно зводять до усної розмовної мови, дістав і текстуальне підтвердження. Його використання в цих текстах, поза сумнівом, було засобом для пропагування й утвердження цього варіанта мови» [11, с.42]. Можна зробити висновок, що цей крок із боку радянської влади сприяв не тільки закріпленню русизмів у структурі української літературної мови, а й додатково розширив межі використання суржику.

4. 1970 – 1980-ті роки ХХ століття. Цей період характеризується посиленням русифікації у всіх сферах. Усе більше й більше людей обирали російську мову як основу спілкування, тому що вона була більш престижною. Через це почався процес руйнування сільських осередків української мови та культури. Процеси урбанізації супроводжувались і розширенням меж використання російсько-українського суржику, оскільки він вийшов за

межі лише окремих осередків. Суржик – наслідок «потужного однобічного впливу російської мови впродовж періоду радянської лінгво-етнічної міксації» [21].

5. 1990-ті – сучасність. Разом із процесом здобуття незалежності України, також змінився і статус української мови. Спочатку в багатьох сферах суспільного життя ще використовувалася російська мова, тому що Конституція Української РСР 1978 р., якою користувались на той час, дозволяла це робити. [20, с. 54]. Далі ставлення до української мови поступово змінювалося і вона поширювалася серед населення. Процес дерусифікації відбувався повільно, а прискорився у 2019 році, коли було підписано закон «Про забезпечення функціонування української мови як державної», який захищає українську мову як державну у всіх публічних сферах суспільного життя: «7. Статус української мови як єдиної державної мови передбачає обов'язковість її використання на всій території України при здійсненні повноважень органами державної влади та органами місцевого самоврядування, а також в інших публічних сферах суспільного життя, які визначені цим Законом» [25]. У тому ж році запроваджено новий правопис, що повернув дуже багато питомо українських мовних особливостей. Він викликав неоднозначну реакцію як серед дослідників, так і серед населення України. Наприклад, Гримашевич Г. подає думку про те, що варіативність форм, зазначена у правописі, може розхитувати мовні норми, але, загалом, робить висновок, що новий правопис «відображає саме українські національні риси» [12, с. 27-28].

Спираючись на різні погляди науковців у питанні походження суржику, можна зробити висновок, що основними причинами появи цього явища є, насамперед, тривалий вплив російської мови на українську, спордненість мов, територіальна близькість держав, а також двомовність та неосвіченість у розрізненні фонетичних, лексичних та граматичних особливостей кожної із мов.

1.3. Суржик та білінгвізм в аспекті інтерференції

Білінгвізм – явище, коли у суспільстві розмовляють двома мовами. У більшості випадків, білінгви використовують обидві мови рівноправно, залежно від певних потреб мовця у конкретній ситуації. Зазвичай, обидві ці мови закріплені як офіційні Конституцією держави. До таких країн належать, наприклад, Канада, Фінляндія та Бельгія. Сам білінгвізм має низку власних ознак: послідовність засвоєння мов (мови поділяють на першу, другу і т.д.), спосіб засвоєння мов (природний, вдома, від носіїв, та штучний, наприклад, у школі), пріоритетність використання (первинна та вторинна мови), рівень володіння (повний та частковий білінгвізм у залежності від рівню знання мов). [35, с. 84-85]

Використання мов білінгвами можна поділити на кілька типів – перемикування кодів, інтерференція та повна інтеграція кодів. Бацевич Ф. подає таке визначення поняття «перемикування кодів»: «Перемикування (переключення) кодів — усвідомлений перехід мовця у процесах мовного (зокрема міжкультурного) спілкування з матричної мови (діалекту, стилю) на іншу мову (діалект, стиль), пов'язаний зі зміною параметрів комунікативного акту» [2, с. 133]. Оскільки цей перехід є усвідомленим, то як приклад можна навести ситуацію, коли іноземець запитує щось англійською. У цей момент людина свідомо обирає використання англійської мови, тому що по-іншому її не зрозуміють. Згадуючи те, що суржик – суміш двох мов, можна сказати, що до цього типу білінгвізму його віднести не можна, оскільки переходу з однієї мови на іншу не спостерігається.

«Інтерференція - взаємодія мовних систем, мовленнєвих механізмів, вплив системи рідної мови на мову, що вивчається, в процесах оволодіння нею. Виявляється у відхиленнях від норм і системи другої мови під впливом рідної» [2, с. 69 – 70]. Це щось середнє між перемикуванням та змішуванням мов. Саме змішування (інтеграція кодів) – процес, коли мовець хаотично використовує засоби мов, якими володіє. Зазвичай, чітка межа між мовами є відсутньою або ж мовець недостатньо володіє однією з мов для вільного висловлювання своєї думки. Суржик найбільш наближений до цього типу поведінки білінгвів. Через

історичні обставини багато українців розуміють російську мову на розмовному рівні. Ситуація навпаки, коли росіянин розуміє українську, є малоімовірною. Саме тому, використовуючи суржик, мовець лексично наближає своє мовлення до російської мови, але зберігає українську основу, зокрема, фонетику, морфологію та синтаксис. [21] Разом із цим, мовна ситуація в Україні є неоднозначною. Є ситуації, коли у мовленні використовуються російські слова у фонетичному звучанні наближені до українських не тому, що людина є носієм суржику, а тому, що вона вивчає українську, до цього володівши лише російською. Таку ситуацію не можна віднести до змішування, оскільки це помилки під час вивчення мови, які можна коригувати. Це відбувається і в процесі вивчення інших мов, коли в особи недостатньо знань про системи мови, яку вони вивчають. Це і є інтерференцією, це нормальне явище на початкових етапах вивчення. З часом, ці помилки зникають, якщо усвідомити мовну межу. Тут є суттєва різниця між білінгвом та суржикомовцем. Двомовна людина має цю межу, вони просто «перемикають» між унормованими варіантами мов, а суржикомовець – ні, тут хаотичне змішування двох мов продовжується [21, с. 57 – 58].

Критерії розрізнення інтерферованого та змішаного мовлення відсутні в українському мовознавстві, що також ускладнює і виокремлення суржику та наведення єдиного чіткого визначення того, що це за явище. Також є відсутніми і ґрунтовні праці, що містили б у собі збірку принаймні найпопулярніших варіацій суржикових слів, за допомогою яких і можна було б встановити такі критерії. Загалом, можна сказати, що білінгвізм є однією із причин витворення суржику, в основному тому, що мови дійсно є близькими між собою. Але, водночас, не можна віднести суржик як явище до класичного поняття білінгвізму.

1.4. Використання суржику в українськомовному соціумі

Суржик – явище, яке доволі глибоко вкорінилося у мовлення українців за останні роки. Зараз дуже важко знайти людей, які хоча б раз не використовували суржикові слова у своєму мовленні. Суржик побутує та поширюється на території України, ненормативні форми закріплюються, а мовці вже не вважають їх неправильними, коли використовують. Але інше питання у тому, що суржик закріпився і письмово – в Інтернеті, у текстах ЗМІ, у художній літературі, у піснях і т.д.

Суржик може використовуватися свідомо або несвідомо. Несвідомо мовці використовують його тоді, коли вважають, що недостатньо знають літературну форму української або російської мов. Деякі мовці вважають, що використання суржику виражає простоту думки. Для інших – його використання є просто зручним, бо таку «мову» чули з дитинства, тому що один або обоє батьків розмовляли суржиком. Свідоме використання суржику зазвичай є відображенням бажання бути «своїм», досягти рівності у статусах: «... суржик стає дієвим засобом комунікативного пристосування мовців один до одного, які керуються прагматичними настановами: «Буду говорити, як ти, щоб досягти кращого порозуміння» [5, с. 124].

Інша ситуація спостерігається, якщо досліджувати вживання суржику у письмовому вигляді, особливо у текстах ЗМІ та художній літературі. Такі тексти зазвичай проходять перевірку перед публікацією і доволі важко «несвідомо» і «ненавмисно» використати суржикіزمі.

За Дзюбишиною-Мельник Н. [15], існує чотири етапи у використанні суржику у художніх текстах:

1. використання суржику у творах XIX - початку XX ст. для створення саркастичного образу задля висміювання манери ведення розмови.

Наприклад, «Такое дело случилось,— отказував Роман.— Как вишел я з служби, зараз должность мене предложили, очинь даже хорошую: триста рублей жалования.» [14].

2. у 20 – 80-тих рр. ХХ ст. суржик використовується не лише як засіб сарказму, а ще й для створення комічності – персонажів усе ще висміюють, але автор уже ставиться до них з позитивом. Наприклад, «Для того з собою взяв: ... д) кошика, де вмістилося б півдесятка тарані, пляшка перцівки ("довоєнного качества"!), / 2 ф. хліба й ін., // загрузка... // е) посвідчення особисте, / є) учотну картку, / ж) удку... [7].
3. у 80 – 90 рр. ХХ ст. суржикомовці поступово займають місце нормальних персонажів, а висміювати починають українськомовних героїв твору.
4. у ХХІ столітті починають створювати тексти, де носієм суржику є освічена та розумна людина. Ці твори є «штучними». Прикладом такого твору є твір Бриниха М. «Шахмати для дибілів»: «Дорога була не очинь далеко — продавщица жила у тому ж кварталі. І верталася всігдa ночью. І ліхтарі на її путі якщо й були, то нікада нормально не світили» [6].

Музика також є невід’ємною частиною життя людини і може доволі сильно впливати на підсвідомість людини. Суржик як явище набув популярності серед співаків, які активно намагалися наблизитися до мовлення населення. Таким чином вони хочуть показати, що можуть поставити себе на місце звичайної людини, яка розмовляє так, як їй зручно. Інша причина, для чого суржик використовується у текстах пісень – висміювання такого мовлення, як і у випадку з літературними творами. Першим гуртом, що використав суржикізм у тексті, був «Брати Гадюкіни», що виступив на конкурсі «Червона Рута» у 1989 році. [4, с. 49-50]

Прояви суржику у пісенних текстах:

- лексичні кальки з російської;

- неправильне використання правил дієслівного керування та родів іменників, прикметників;
- недотримання орфоепічної мовної норми [32, с. 40].

Найбільша проблема використання суржику як стилістичного або мовного елемента у художніх та пісенних творах полягає у тому, що це, по факту, надає людям упевненості в тому, що їх повсякденне мовлення є цілком нормальним, адже це ж письменники, автори пісенних текстів, вони повинні бути освіченими.

Та сама ситуація спостерігається і з використанням суржику у текстах ЗМІ. На початку ХХІ століття в мові українських медіа активно почали використовувати суржик.

Загалом, у журналістських текстах суржик виконує такі функції:

- номінативну (журналіст свідомо використовує русизм або просто не має правильного відповідника);
- експресивно-оцінну (для висміювання чужого мовлення);
- відтворення мовної характеристики (для передачі чужої мови) [19, с. 108].

Суржик на радіо також використовують з певною ціллю. Він, як і у попередніх випадках, є елементом створення думки, що мовець не один і що цілком нормально використовувати таке мовлення в нормальному житті. [23]

Загалом, можна зробити висновок, що не в побутовій сфері суржик використовують в основному з двох причин – бажання наблизитися до аудиторії, відобразити стан та мовлення суспільства або ж для створення ефекту комічності.

Інший бік – використання суржику в Інтернеті. Його можна віднести до побутового суржику, оскільки носіями є звичайні користувачі соціальних мереж. Шевчук-Клюжева О. у статті «Суржик в Інтернет-просторі (на матеріалі соціальних

мереж Вконтакте, Facebook, Instagram)» навела як приклад популярну сторінку «Чоткий паца». Там активно оновлювався контент, який, зазвичай, містив суржикізми. Наприклад, «Новий Год по погоді все більше наминає Паску». У цьому випадку мета використання суржику – це створення жартів. Але це породжує примітивізацію української мови, викликає враження, що українська мова може використовуватися лише для таких висміювань. [39]

В основному суржик є явищем усного міжособистісного спілкування. Хтось засвоїв його від батьків, хтось від оточення, є люди, яким простіше спілкуватися суржилом, а є й такі, які використовують його лише заради жарту. Приклад такої розмови:

« – Іду варить макарони і катлету їсти.

– А в мене 2 пачки пельменів є.

– Блін, повезло тобі, тоже хочу пельмені їсти» (приклад взято із особистого спілкування).

У цьому випадку обидва носії можуть спілкуватися як українською, так і російською мовами (один з них переважно спілкується українською, а інший – російською), але все-таки обирають щось середнє – суржик.

Вважаємо, що питання вживання суржику у різних сферах суспільного життя не є вичерпаним і заслуговує детальнішого вивчення, адже можливі різні ситуації – хтось вживає суржик свідомо, а хтось – несвідомо як наслідок двомовності суспільства та затирання різниці між двома мовами.

1.5. Лінгвістичні ознаки суржику

1.5.1. Типізація суржикового мовлення

Суржик як мовне явище не є унормованим, отже, його типізація є складною. Дослідники поділяють суржик на доволі різні типи, наприклад «на основі української мови», «слабкий», «сильний», «сільський», «міський», «радянсько-український» та інші. Зазвичай, типи виділяють або за ступенем порушень норм літературної мови, або за класифікацією, або за осередками вживання. Також суржик відрізняється за регіонами, тому що у людей із різних областей різний рівень засвоєння російської, а також специфічний діалект української мови, що призводить до ще більшої проблеми, оскільки охопити усі наявні форми одного слова не є можливим.

Але, насамперед, можливо виділити певні риси, за якими слово можна віднести до суржику. Наприклад, найбільш очевидним є вживання русизмів замість українських слів-відповідників. Зазвичай, вимова таких слів наближається до українських норм: *делать* – [д'`ела-т'], *лішня* – [л'`ішн'а], *короче* – [к^oр^oоче] *тощо*. Щодо дієслів – є два варіанти: або використовують власне українське дієслово-інфінітив із закінченням –ть: *гулять* (*гуляти*), *бігать* (*бігати*), *казать* (*казати*), або російські дієслова, підлаштовані під вимову та граматику української мови: *устроїться* – [устр^oоїц'а], *учітувать* – [у-ч'`ім^oува-т'], *попробувать* – [п^oопр^oоб^oува-т']. Типовим також є змішування форм для займенників: *якій-то* (*якийсь*), *шо-то* (*щось*), *коє-шо* (*дещо*), *чим-то* (*чимось*), *хто-то* (*хтось*). Утворення ступенів порівняння за російським зразком дуже сильно проникло в українську мову, тому такі помилки описані навіть у підручниках: *саме лучше* (*найкраще*), *саме сложне* (*найскладніше*), *самий безболєзненний* (*найменш болісний*). [17, с. 50 - 53]

Синтаксис у суржику теж страждає. Найтиповішим є калькування російських форм дієслівного керування, вживання прийменників та відмінків. Наприклад,

ключом (ключем), в Фейсбуке (на Фейсбуці), на англійській (англійською). Поширеними є слова та вирази, які не є нормативними в українській мові: *прийняти участь, на даний момент, на протязі тощо.*

Вимова слів зазнає найбільш значних та відчутних змін. Перше, що варто згадати, - це пом'якшення усіх приголосних звуків у словах. У російській мові більшість приголосних у оточенні інших звуків стають м'якими, українська вимова слів цим дуже відрізняється від російської: *треніровка (тренування), волейбой (волейбол), пітаніє (харчування), ощуценіє (відчуття).* Також можна відмітити редукцію ненаголошених голосних, оглушення дзвінких приголосних, тощо.

Ці риси можна віднести до тих, за якими зазвичай розпізнають носія суржикового мовлення. Можна навести також інші, але найбільш помітними є використання російських словоформ та калькування вимови слів.

Аналіз суржика та його типізація залежить від наявності матеріалів для цього: текстів, що представляють суржик у різних контекстах. Така база даних повинна бути доволі великою і різноматичною для повного дослідження вживаних суржикових словоформ. На жаль, такої бази даних не існує. Все, що ми маємо – окремі слова чи фрази, що зустрічаються художніх, журналістських, політичних та соціальних працях, а також ніяк не зафіксоване усне мовлення населення [41, с. 97]. Нижче подано типи нестандартних мовних явищ на різних рівнях, які можуть бути включені у поняття суржик.

1.5.2. Фонетичні ознаки суржика

Як було зазначено раніше, фонетична система суржика має більш близький зв'язок з українською мовою, ніж з російською. У більшості випадків у суржиковому мовленні відбувається адаптація фонетики російських слів до орфоепічних норм української мови. Багато українців говорять мовними

варіаціями, які відображають норми літературної мови, але використовують при цьому фонетичні особливості зовсім іншої мови [41, с.98].

Частотним є явище, коли люди плутають вимову двох звуків, які є спорідненими, але по-різному артикуються. До них належать український голосний звук [и] та російський [ы]. У російській мові звук [ы] артикулюється глибше в роті, що робить його твердішим, тоді як український звук [и] є менш твердим.

Російське фонологічне правило «акання», коли ненаголошений голосний [о] вимовляється як [а], поширюється на суржикове мовлення: *пагано* – [пaг`aн^oo], *кароче* – [кaр^oочe]. Але також можливе і збереження відсутності редукції цього голосного, наприклад, *человек* – [чeл^oo·в·`eк].

Більшість слів у російській та українській мовах відрізняються замінами літер та звуків, які виникли у процесі історичного розвитку мов. В українській мові, наприклад, використовується звук [і], який є відповідним російським звукам [о], [е] у спільнокорених словах (наприклад, *"лес"* українською звучить як *"ліс"*, а *"род"* – як *"рід"*). Українська мова зберігає твердість приголосних звуків перед [и] та [е], тоді як у російській мові ці приголосні стають м'якими. У суржику можливе збереження цього правила для російської лексики, але у більшості випадків ці слова звучать за російськими правилами, але пишуться українськими літерами, наприклад, *серйозно* – [с·`eрj^o·`oзн^oo], *реально* – [р·`e`a·л·н^oo]. [24, с. 49]

Дивною є і ситуація з приголосними звуками. У випадку етимологічного задньоязикового звуку [g] складна ситуація стає ще більш складною. Згідно з дослідженнями, цей звук є природнім для української мови і виник як результат одзвінчення [к] після падіння зредукованих. Існування фонем [h] і [g] відокремлює українську мову від російської та південнослов'янських мов, і наближає її до романських та германських мов, де також існують ці фонemi. Російськомовні люди вимовляють свій звук [h] як [g]. Фарингального щілинного

приголосного [h] немає в російській мові (цей звук може з'являтися у деяких лексемах, таких як вставні слова, наприклад, *ага*). Українська мова має літеру г для позначення задньоязикового проривного звуку [g] та літеру ґ для позначення [h]. Багато людей, що розмовляють російською, схильні говорити /*год*/ замість нормативного російського /*гом*/ (українською рік). Або ж, можлива ситуація навпаки: у місці, де за нормами української мови повинен бути фарингальний щілинний приголосний [h], хтось може використовувати задньоязиковий проривний звук [g]. У суржику [h] уживається як в словах, запозичених з російської, так і в словах, де, згідно новим правилам, повинен вимовлятися задньоязиковий [g] (наприклад, *грунт*) [41, с.98].

В українській мові приголосні [ч] і [щ] за орфоепічними нормами є твердими. Існує лише декілька винятків з цього правила, за яких ці приголосні можуть пом'якшуватись (наприклад, слово *обличчя*). Це правило найчастіше збережене у суржиковій лексиці, що походить із російської мови (*вообще* – [в^о:биц`е]). Але в той же час можливий і випадок, коли ці приголосні стають м'якими.

Під впливом російської мови спостерігається заміна африкат [дз] і [дж] на приголосні [з] і [ж] у деяких випадках, коли в російських словах зустрічаються ті самі звуки у тій же позиції: наприклад, *звонять, позвонив* (замість *дзвонять*). Помітною рисою суржикових приголосних є оглушення дзвінких приголосних (як у російській мові), тоді як в українській мові ці звуки є чіткими та дзвінкими (*розписати* – *расписать*). [1, с. 113, 115]

Хаотичне використання орфоепічних норм двох мов у суржику призводить до того, що одному нормативному слову може відповідати кілька суржикових. Як приклад можна взяти російське слово «*короче*». Тут присутнє явище «акання» і це слово правильно вимовляється як /*кароч`е*/. У суржиковому мовленні можна знайти варіанти /*кароче*/, /*кароч`е*/, /*короче*/, /*кароче*/ і навіть /*коротше*/. Це одна із причин, чому унормувати та стандартизувати явище суржику не є можливим

– занадто багато варіантів вимови та, відповідно, написання слів. Усі варіації відтворити неможливо, адже мовців дуже багато і навіть вони у різних випадках можуть використовувати декілька варіантів вимови та написання.

1.5.3. Лексичні особливості суржику

Так, як і у вимові, так і у лексиці, суржик дуже добре проявляється. Найчастіший випадок – заміна нормативного українського слова російським, як уже було згадано вище. Наприклад, *нравиться (подобається), на pewno (напевно), лічно (особисто), парня (хлопця), родителей (батьків) тощо*. Інша лексична сторона суржику – слова, що пишуться і звучать однаково, але мають різне значення. До таких слів належать, наприклад, слова: *час (година) і час (час), неділя (день тижня) і неделя (тиждень)*. Їх використання як взаємозамінних може викликати втрату змісту у реченні. У деяких випадках українські терміни втрачають можливі значення, що відрізнялися від російських. Подібним прикладом є вживання українського робити в значенні «працювати». У стандартній українській мові це означає «робити», а «працювати» - це практичність. Однак під впливом російського «работать» спостерігається лексичний зсув серед деяких носіїв суржику, коли «робити» означає те саме, що і «працювати» [41, с.98]. Інший приклад – використання слів у невласливих їм значеннях. Вислів «я рахую» є калькою з російської мови, де слово «считать» має 2 значення: «рахувати» та «вважати». В українській мові це значення тільки одне – «рахувати». Багато неологізмів, жаргонізмів та розмовної лексики переносяться у суржикове мовлення із російської (*общага, ору*) [22, с. 71]

Загалом, за Мозером М., у суржику відбуваються:

- запозичення лексичних одиниць – слово запозичується з російської мови в його нормативній формі (*арбуз, говядина, головка, горсть, градусник, мужчина, ноль, мусор*);

- розширення значень лексичних одиниць – слово присутнє і в українській, і в російській мовах, але має різні значення; саме за рахунок значення із російської мови відбувається таке розширення (*час, год, город*);
- лексичні кальки - слово або вираз, скопійовані засобами рідної мови з іншої мови; кожна значуща частина слова-оригіналу буквально перекладається і займає в перекладі таке ж місце, як і в оригіналі (*накінець, багаточислений, безперервно, больниця, бумага, зачот, зонт, туалет*). [24, с.34]

За Фльором М., морфеми у різних частинах мовах підлягають або не підлягають русифікації:

- у займенниках жодні частини не можуть змінюватися;
- у дієсловах можуть бути російськими тільки корені;
- в іменниках тільки префікси не можуть бути російськими тощо; [42]

Закриті класи морфем частіше за все залишаються українськими. Але, можемо навести приклад, коли морфеми займенників змінюються – *хто-то* замість *хтось* або *де-то* замість *десь*.

Варто також відмітити, що на лексичні зміни накладаються і фонетичні, що призводить до витворення достатньо дивних і хаотичних форм і, як уже згадувалося раніше, це призводить до існування паралельних форм одного суржикізму.

1.5.3. Синтаксичні особливості суржикі

Синтаксичні форми однієї мови можуть використовуватися в іншій. Поширеним «русизмом» в українській мові є використання місцевого відмінка замість орудного, коли кажуть, наприклад, «*Дивлюсь відео на іспанській мові*» (калька із російської), а не «*Дивлюся відео іспанською мовою*». Той факт, що російське

слово для мови – язык – є чоловічого роду, тоді як український термін мова є жіночого, часто призводить до додаткового неправильного використання унормованих форм. Це явище розповсюджується і на інші слова, що мають різний рід в українській та російській мовах.

Грамматика слів, словосполучень та речень теж зазнає суттєвих змін. Рід, граматичні та відмінкові форми піддаються калькуванню, що призводить до неправильної побудови словосполучень та речень. До прикладу, після числівника «два» в українській мові повинен йти іменник у множині: «два *рази*». У російській мові це словосполучення прозвучить як «два *раза*». Форма «*раза*» у цьому випадку є характерною для мовлення носіїв суржику. Крім простих запозичень, спостерігається також і використання російських словотвірних елементів для утворення слів. Наприклад, *спать (спати), фотографіровать (фотографувати), робить (робити) тощо*. [22, с.71]

Використання прийменників також не є унормованим. Грамматичні правила їх вживання часто відрізняються і там, де не повинно бути прийменника в українській мові, він може використовуватися в російській мові. Наприклад, «*план по завоюванню*», а не «*план завоювання*». Можлива і ситуація навпаки. У словосполученні можуть використовувати іншу відмінкову форму або інший прийменник (*ходить по магазинам – ходити по магазинах, із-за нього – через нього*). [22, с.73]

Зазвичай, речення у суржику будують відповідно до правил як і української, так і російської мов. Як наслідок – відсутність меж між конструкціями нормативно українськими та нормативно російськими.

Висновок до розділу I

Розділ I присвячено теоретичним засадам трактування явища суржику у лінгвістиці. Детально розглянуті питання визначення суржику як явища,

передумов та причин його виникнення, вживання у різних сферах, а також характерні особливості та різні типи суржику.

У лінгвістиці існують різні погляди на те, що таке суржик та яке місце він займає у мовній системі. Через це виникає також багато інших проблем – зокрема, проблема типізації суржику. Зазначимо, що базовим для дослідження було обрано таке визначення: «Суржик – мовне явище, якому характерне ненормоване змішування фонетичних, лексичних та граматичних особливостей російської та української мов».

Серед причин появи та поширення суржику основними, на нашу думку, є панування Росії над Україною впродовж століть, що призвело до стирання меж між мовами; схожість мов між собою, їх контакт протягом тривалого періоду часу; недостатня обізнаність у якійсь або обох мовах, що призводить до хаотичного використання мовних засобів; установлення російської мови як основної для публічних сфер життя. Саме контактування мов між собою та їх схожість на фонетичному, лексичному та граматичному рівнях, а також таке явище, як білінгвізм, викликали змішування елементів та структур української та російської мов. Вперше суржик з'явився у кінці XVII століття та поширився серед українського чиновництва та солдатства. Варто відмітити, що найважливішою причиною закріплення суржику стала політика зближення мов у 1930-х рр. , через яку багато питомо українських слів замінили російськими відповідниками.

Суржик активно використовується не лише у побуті та усному мовленні населення, а й у художній літературі, піснях, текстах ЗМІ тощо. Загалом, це призводить до нормалізації цього явища у суспільстві, оскільки люди зазвичай вважають, що у книжках та по телевізору усі мовні засоби вживаються правильно.

РОЗДІЛ II. КОРПУС СУРЖИКОВИХ ТЕКСТІВ

Корпусна лінгвістика – галузь прикладного мовознавства, яка зосереджена на створенні, обробці та аналізі великих масивів комп'ютерних текстів, які становлять корпус текстів. МакЕнтрі Т. зазначає, що корпусна лінгвістика відрізняється від інших галузей лінгвістики, таких, як, наприклад, синтаксис. Синтаксис зосереджений на поясненні того, як мовці будують речення, тоді як корпусна лінгвістика – це методологія для подальшого дослідження [46, с. 2]. Але більш поширеною є думка про те, що корпусна лінгвістика – самостійна наука, яка має свої власні характеристики для того, щоб такою вважатися.

Розвиток комп'ютерної лінгвістики почався у 60-х рр. ХХ ст. через швидке поширення комп'ютерних технологій. Але термін закріпився лише у 1990-х роках із виходом збірки «Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research» за матеріалами конференції ICAME «Conference on the Use of Computer Corpora in English Language Research» [16, с. 8]. Корпусна лінгвістика має неоднозначне становище як наука. З одного боку, багато дослідників вважають, що поява корпусів значно полегшила дослідження, тому що тепер усі матеріали можна зберігати у комп'ютерному форматі, що допомагає швидше фільтрувати та шукати необхідну інформацію, а також укладати різні словники за матеріалами корпусу. Комп'ютерні корпуси дійсно ефективні, коли потрібно опрацювати великий масив текстів. Але з іншого боку, є твердження і про те, що комп'ютерна лінгвістика і власне корпус не є хорошою ідеєю для дослідження. Зокрема, Хомський Н. був противником методів корпусної лінгвістики, оскільки вважав, що корпуси відображають лише окремі зовнішні висловлювання і ніяк не відображають мовну компетенцію, яка складається із знання мови та їх застосуванні на практиці [10, с. 250]. Також частково поширена думка, що корпус не повинен бути єдиним джерелом та методологією досліджень, його треба поєднувати з чимось іншим.

Щодо того, що ж вважати корпусом також є різні думки. Дехто вважає, що він не обов'язково створюється з якоюсь конкретною метою, інші – що він створюється суто з конкретною ціллю і відповідно має бути строго систематизованим та розміченим [18, с. 236]. Загалом, корпус можна визначити як масив даних, що поданий в електронному форматі, правильно структурований, розмічений та компетентний для представлення якихось характеристик мови.

Існують різні типи та класифікації корпусів. Зокрема, Кеннеді Г. вважає, що корпуси слід поділяти на не електронні/електронні, загальні/спеціалізовані, усномовні/письмові, тренувальні/тестувальні, діалектні тощо [44]. У інших класифікаціях можна знайти такі поняття як простий/анотований, одномовний/двомовний/багатомовний, динамічний/статистичний корпуси. Загалом, класифікація проводиться за двома основними способами: протиставленням національних корпусів спеціальним та за типом лінгвістичної розмітки.

Для розмітки корпусів не існує якихось певних обмежених стандартних вимог. Дослідники, укладаючи їх, орієнтуються на своє бачення та необхідність певної інформації для дослідження, але, вважається, що чим краще анотовано корпус, тим легше потім ним оперувати. Існують метатекстова, структурна та лінгвістична розмітки. До метатекстової входить все, що стосується загальної інформації про текст, наприклад, його автор. Структурна розмітка зосереджена на відокремленні окремих елементів у текстах, наприклад, речень та слів, а лінгвістична – приписуванні окремим елементам певної лінгвістичної інформації.

Оскільки для нашого дослідження проводилася робота із записами, можна встановити, що корпус (початково) є усномовним. Корпус усного мовлення – спеціалізований корпус текстів (окремих його уривків, речень, словосполучень, слів тощо), вимовлених різними мовцями у різних умовах для подальшого

дослідження. Робота з усномовними корпусами є трудомісткою, адже для початку потрібно опрацювати усі отримані записи. Загалом, серед етапів роботи із усномовними корпусами виділяють власне процес запису (збирання даних), транскрибування та розмітку отриманих аудіозаписів, їх упорядкування та сам аналіз [40, с. 40]. Такі корпуси важливі для отримання різної інформації, наприклад, про вимову слів та звуків у контексті, а також масиву опрацьованих аудіозаписів, що буде корисним для розробки та розвитку систем розпізнавання та синтезу мовлення.

Корпус суржикових текстів є, насамперед, корисним для систем розпізнавання українського мовлення. Зазвичай, такі корпуси натреновані на розпізнавання літературної української мови із врахуванням певних особливостей вимови. Використання суржику мовцями у різних записах знижує відсоток успішного розпізнавання, оскільки суржикізм вживаються хаотично та непослідовно і не входять у вибірку нормативної української лексики. Корпус суржикових текстів є корисним і для проведення інших досліджень. Наприклад, словник «суржикізм – нормативний варіант» можна використати для розроблення програми для виправлення помилок у текстах. Транскрипції суржику у поєднанні із прив'язкою до аудіозаписів дають уявлення про специфіку вимови суржикових слів різними носіями; за допомогою словнику словоформ можна обчислити частоту вживання окремої словоформи у всіх текстах або у окремих текстах; весь зібраний словник лексем є корисним для встановлення статистичних даних щодо загального відсотку суржикізмів у всьому масиві текстів. Також окремою задачею є розробка інтерфейсу для отриманого корпусу, що дозволить звичайним користувачам послухати вимовлені у контексті суржикові словоформи або ж знайти відповідник до конкретної суржикової словоформи.

2.1. Етапи укладання корпусу суржикових текстів

2.1.1. Збір аудіозаписів

Джерелом мовленнєвого матеріалу для проведення дослідження є аудіозаписи усного мовлення, що містять суржикізми. Записи, загалом, є уривками із розмов у соціальній мережі Telegram. Кожен аудіозапис є зразком усного спонтанного мовлення на побутову тематику, що включає обговорення навчання, щоденних справ, вираження думок щодо різних речей тощо. Записи не були спеціально створеними для проведення дослідження, оскільки важко говорити про репрезентативність суржикового мовлення, якщо людина свідомо знає, що їй треба сказати щось із суржигом. На нашу думку, саме такі записи є показовими для дослідження суржику як мовного явища, оскільки диктори не були зосереджені на виправленні свого мовлення або підборі правильного варіанту слова. Усі зібрані аудіозаписи є частинами особистих розмов у період із 2019 до 2022 року (присутній також один аудіозапис 2023 року).

Отож, для дослідження відібрано 135 аудіозаписів формату .wav різної тривалості — від 15 секунд до 5 хвилин. Загальна тривалість усіх оброблених аудіозаписів становить годину сорок хвилин та сорок секунд, тобто, приблизно 6041 секунду усного мовлення. Дикторками стали три респондентки, усі жіночої статі та у загальній віковій категорії від 14 до 18 років у різні роки (наприклад, одній дикторці у 2019 році було 14 років, а у 2022, відповідно, 17 років). Усі три респондентки родом із Києва та Київської області. За наявності більшої кількості дикторів із інших регіонів можливо також провести дослідження для виизначення регіональних особливостей використання суржикової лексики та відмінностей у вимові.

Після прослуховування та відбору запису, він вивантажувався із Telegram у форматі .ogg та переформатовувався для зручнішого та повнішого дослідження у формат .wav. Кожен із записів отримав власну назву, яка відповідає порядковій

цифрі його опрацювання – від 1 до 135 відповідно. Усі аудіозаписи можна переглянути та прослухати у папці за посиланням у [Додатку 2](#).

2.1.2. Розпізнавання усного мовлення

Для подальшої роботи із аудіозаписами потрібно було провести розпізнавання усного мовлення. Цей процес є доволі важким, оскільки аудіозаписи не завжди чіткі, на фоні присутні шуми, що можуть впливати на якість розпізнавання. Оскільки аудіозаписи отримані із соціальної мережі Telegram, то початкове розпізнавання аудіозаписів проводилося за допомогою вбудованої функції автоматичного розпізнавання мовлення саме там. Воно не є досконалим, адже автоматичне розпізнавання усного мовлення є важкою задачею, а наявність суржикової лексики ще більше його ускладнює.

Як приклад наведемо початковий варіант розпізнаного тексту для [аудіозапису 119](#): *«Ну в общем типа Ось я уже почти 11:30 я до их портягую время того чтобы ты спать ну начался потому что я хотела поиграть в игру часи потому что я хотела кушать вот частью Потому что я не хочу спать но морально тоже чувствую что я якость мене щось стримуе ити спати а що я не знаю От такі діла»*. Цей варіант не є точним відображенням сказаного мовцем, у ньому недосконало розпізнаний суржик (це, зазвичай, російський нормативний варіант слова або зовсім інше слово), а також відсутні розділові знаки. Для систематизації та опрацювання текстів та його елементів, зокрема, суржикізмів, цей варіант не підходить. Тому, редагування розпізнаного мовлення проводилося вручну безпосередньо при створенні текстгридів для записів у програмі Praat.

Кінцевий варіант розпізнаного тексту для цього ж аудіозапису виглядає так: *«Ну, в общем, типа, ось уже почти одинадцять, без пяти одинадцять, я до six пор оттягую время того, щоб ити спать, ну, от часті потому, що я хотіла поіграть в ігру, от часті потому, що я хотіла кушать, от часті потому, що я не хочу спать, но морально тоже чувствую, що я якость, мене щось стримуе ити спать,*

а що, я не знаю. От такі діла». Цей варіант є текстовим представлення сказаного в аудіозаписі. Варто зазначити, що розділові знаки при цьому розставлялися не настільки детально, оскільки для спонтанного мовлення важко визначити межі сказаного, особливо кінець речення та початок нового. Але у роботі використані такі розділові знаки:

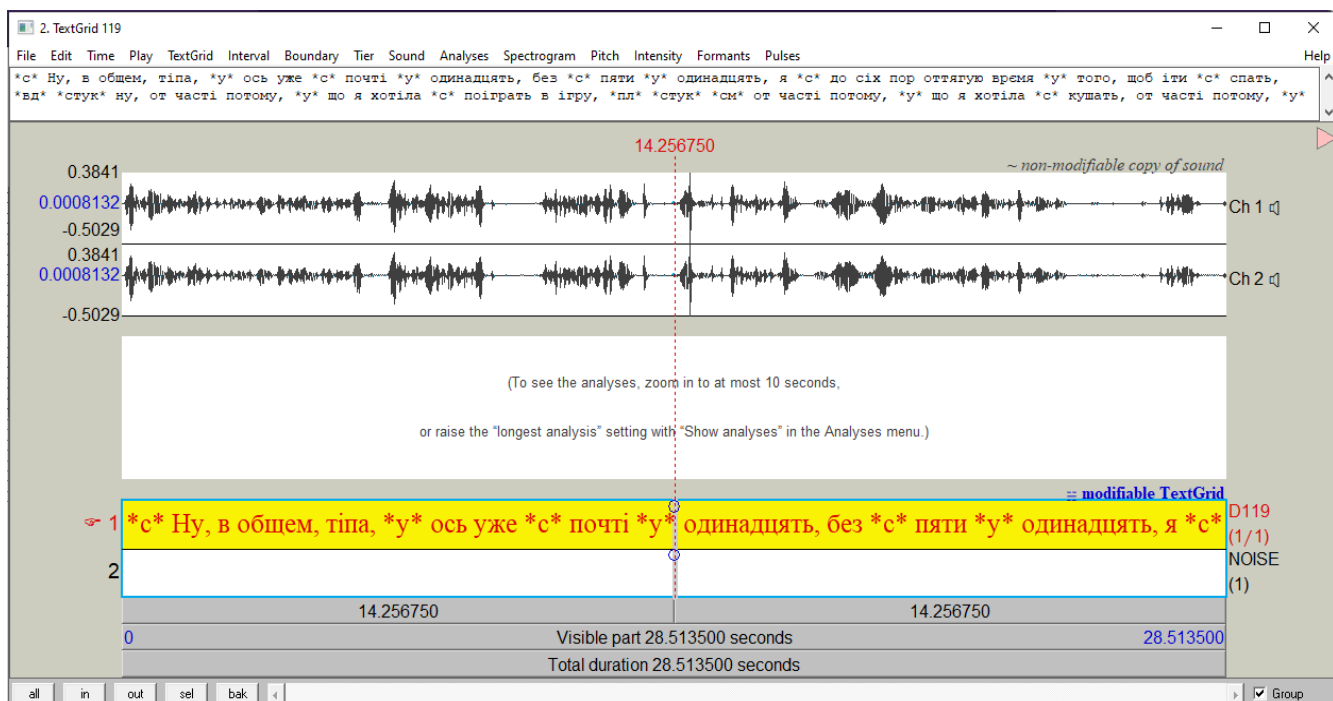
- *крапка*: синтагма має завершальну інтонацію, далі пауза, кінець речення;
- *кома*: позначає незавершеність синтагми або інші моменти, що залежать від граматичних правил (напр., кома перед "а");
- *тире*: вживається відповідно до граматичних правил;
- *знак питання і оклику*: синтагма має питальний або окличний характер;
- *двокрапка*: вживається відповідно до граматичних правил;
- *крапки*: позначають очевидно незавершене або обірване речення;
- *лапки*: використані для власних назвах, дослівної цитата або лапки "в повітрі".

Інші розділові знаки також вживаються відповідно до граматичних правил та доречності їх використання. Велика літера використана на початку речення, для імен та власних назв тощо.

2.1.3. Обробка записів у програмі Praat та створення TextGrid

Praat – комп’ютерна програма, за допомогою якої можна аналізувати, синтезувати та досліджувати мовлення із різних поглядів. За допомогою неї можна створювати TextGrid – спеціальний файл, у якому зберігається текст розпізнаного мовлення, відбувається його розбиття на синтагми (за потреби), а також додаються різного роду примітки експерта, які той вважає потрібними для своєї роботи. Простіше, текстгрід – анотація сказаного мовцем у записі. Для всіх 135 аудіозаписів створено 135 текстгрідів, які мають таку ж назву, як і опрацьований аудіозапис, та збережені у форматі .TextGrid. У текстгрідах для кожного аудіозапису міститься два рівні – рівень Dn, де n – порядковий номер

запису, та рівень NOISE. На рівні Dn відтворено у письмовій формі мовлення диктора (процес його розпізнавання згадано у попередньому підрозділі), до якого додані позначення про хезитації, шуми та мову диктора. Рівень NOISE наявний для позначення шумів, якщо вони присутні у всьому записі.



Малюнок 2.1. Аудіозапис 119 та TextGrid 119 у програмі Praat

Як уже було згадано, у текстгріді, крім самого тексту, містять також і позначки, що стосуються різних мовних та позамовних проявів. Вони розташовані прямо у тексті та позначені різними скорочення у поданні *n*, де n – скорочення.

Список позначень, використаних у роботі:

- позначення мови: відноситься до всіх слів синтагми до наступного нового позначення мови (ставиться перед першим словом відповідної мови):
 - *у* – українська мова;
 - *с* – суржик;
 - *р* – російська мова;
 - *а* – англійська мова;
- позначення способу вимови слів (ставиться перед відповідним словом):

- **о** – обрив; слово обривається, диктор не договорило його (пишеться лише частина слова, яка реально звучить);
- **об** – обмовка; диктор помиляється у вимові й сам себе виправляє (пишеться лише та частина слова, яка реально звучить);
- **роз** – розтягування; слово вимовляється зі значним розтягуванням одного (кількох) звуків (пишеться нормативний варіант слова);
- **пскл** – по-складова вимова; слово вимовляється по-складово (пишеться нормативний варіант слова);
- позначення фону (тла) запису: стосується всього сегменту, ставиться на його початку, відноситься до тла монологу диктора:
 - **шт** – шум на тлі; будь-який шум, на фоні якого говорить диктор;
- позначення хезитацій/стилю мовлення: стосується конкретних звуків у монолозі диктора, ставиться на місці відповідного звуку:
 - **см** – сміх;
 - **наспів** – наспівування;
 - **задих** – задишка;
 - **к** – кашель;
 - **вд** – вдих/видих, втягування повітря;
 - **пл** – плямкання;
 - **шм** – шмигання носом;
 - **е** – екання/акання;
 - **угу** – угукання, агакання;
 - **позіх** – позіхання;
 - **хм** – хмикання;
 - **гкхм** – прочищення горла;
 - **зд** – здивування;
 - **м** – ммм (для позначення усвідомлення, задуманості);
 - **сьорб** – сьорбання;
 - **ковт** – ковтання;
 - **м-м** – вагання;
 - **ім** – імітація звуку;
 - **виг** – вигук;
 - **цьом** – звук поцілунку;
 - **шепіт** – шепіт (диктор шепоче фразу);
 - **сміх** – сміх чи усмішка;

- позначення шуму/неробочих сегментів: позначення стосується всього сегменту чи конкретного моменту, де можна визначити джерело шуму:
 - **шум** – невизначений шум;
 - **скр** – скрип;
 - **шух** – шарудіння;
 - **стук** – стукіт;
 - **маш** – шум машини;
 - **гав** – говоріння;
 - **віт** – вітер;
 - **пік** – пікання (напр., мікрохвильової печі);
 - **птах** – спів птахів;
 - **спов** – звук сповіщення на телефоні;
 - **вібр** – звук вібрації на телефоні;
 - **соб** – гавкіт собаки;
 - **муз** – музика;
 - **нерозб** – диктор говорить нерозбірливо (пишеться замість слів).

Загалом, усі позначки необхідні для того, щоб пізніше використовувати текстґриди для навчання систем розпізнавання мовлення. Найсуттєвішими для подальшої роботи є позначки, що стосуються мови диктора. Зокрема, у записах були присутні 3 мови та суржик: **у** - українська, **р** - російська, **а** - англійська та **с** - суржикове мовлення. Саме завдяки цим позначенням пізніше виділялися суржикові словоформи для опрацювання у корпусі. Готові текстґриди збережено як текстові файли та зібрано у папку, яку можна переглянути за покликанням у [Додатку 3](#).

2.1.4. Текстова представлення аудіозаписів

Тексти, наведені до аудіозаписів, відображають сказане мовцями. Їх отримано із вже опрацьованих текстґридів з попередніх етапів, що стосуються процесу розпізнавання мовлення. Але для текстового представлення вони були очищені від спеціальних позначок, оскільки вони не є суттєвими у звичайних файлах із текстами. Переглянути кожен текст можна за посиланням у [Додатку 4](#). Файли

мають назви від 1 до 135 та збережені у форматі .docx. Кожен із файлів містить текст конкретного аудіозапису.

Саме текстове представлення аудіозаписів потрібне для прив'язки конкретного суржикізму до тексту, з якого його взято. Це допомагає структурувати корпус, а також подати не тільки аудіо-версію вимовленого, а ще й письмову для показового прикладу контексту вживання. Також тексти необхідні для визначення обсягу корпусу, який є суттєвим елементом для подальших досліджень, наприклад, для визначення частоти використання суржику серед усіх текстів.

Загалом, обсяг усіх 135 текстів становить 12 644 слововживань. Найменша кількість слів у одному тексті – 25 у Тексті 67, найбільша – 606 у Тексті 14. Середня кількість слів на один текст дорівнює 93,66 слововживань.

2.1.5. Формування реєстру суржикізмів

Оскільки у текстгрідах до записів присутні примітки про належність словоформи до якоїсь конкретної мови або ж саме суржику, то реєстр суржикізмів було зручно формувати, використовуючи їх, а не текстові файли. Це пришвидшувало роботу, оскільки не доводилося заново перечитувати текст і встановлювати, яке слово є суржикізмом.

Наприклад, маємо уривок із [текстгріду 8](#) для [аудіозапису 8](#): «*у* І то один *с* куда-то *у* уїде, а ви вдвох будете *с* де-то вмєстє, *у* в одному *с* городі, допустім. *у* От чисто отак от і *с* проісходить. *вд* *у* І просто *о* це. *м-м* Я можу сказати, що *с* процентов *у* на 90 або 80, *у* якщо нічого не *с* поміняється, уверєна, *пл* *у* що ми *с* послі *у* школи *с* общаться *у* з ними не будемо». У цьому випадку важливими є лише ті примітки, що стосуються мови, зокрема помітка *с*, що означає вживання суржикового слова. Отож, із шматку цього тексту виділено такий список суржикізмів: *куда-то, де-*

то, вмєстє, городі, допустім, проісходить, процентов, поміняється, увєрена, послі, общається.

Таким чином із кожного тексту ми отримували список суржикових словоформ, які вносилися до корпусу. Варто відмітити, що всі словоформи відтворювалися на слух – тобто, «як чуємо, так і пишемо», оскільки іншого варіанту немає. Це зумовлене роботою із спонтанними усномовними записами.

2.1.6. Лінгвістичне опрацювання суржикізмів

Перш за все, варто вказати, що у самому корпусі суржикізм прив'язаний до запису та тексту, тобто, для кожної словоформи у сусідніх колонках можна знайти та прослухати аудіозапис та почитати його текстове представлення. Що стосується саме лінгвістичного опрацювання, то до кожної словоформи наведена її лексема, нормативний відповідник, транскрипція, розмежування омонімії, тип суржику та інші примітки.

Лексема. Лексема – початкова форма для конкретної словоформи. Для більшості словоформ у корпусі наведено початкову форму, виняток становлять лише певні словосполучення або синтаксичні конструкції, що при зміні втратять свій зміст (наприклад, *не при чому – ні до чого, по крайнєй мерє – принаймні*).

Нормативний відповідник. Відповідник у цьому випадку – це літературний варіант слова із нормативної української мови; з іншого погляду – переклад суржикового слова українською мовою. Він підібраний для суржикової словоформи та наведений у тому ж значенні та відмінку, у якому використаний у тексті.

Транскрипція. До усіх словоформ у корпусі наявні транскрипції, які відображають їх звучання у записі, контексті, а також загалом. Оскільки суржик – суміш двох мов, то вимова також має особливості як української, так і

російської мов. Одні мовці можуть вимовляти слова ближче до російського варіанта, інші – до українського.

Омонімія, тип суржику, примітки. Розмежування омонімії є важливим, оскільки бувають ситуації, коли словоформа є одною і тією ж самою, а значення (лексичне або граматичне) у них відрізняються. Типів суржику виділено три – лексичний, фонетичний та синтаксичний. Примітки стосуються різної додаткової інформації – укавання власної назви, імені тощо.

2.1.7. Укладання бд корпусу суржику та її структура

Використовуючи усі матеріали, створено базу даних корпусу суржику. Корпус створено у вигляді таблиці, де містяться такі колонки:

- словоформа;
- лексема;
- відповідник;
- транскрипція;
- омонімія;
- тип суржику;
- аудіозапис;
- інформація про мовця;
- текст;
- примітки.

Загалом, обсяг корпусу становить 12 644 слововживання із 135 текстів. Фрагмент бази даних представлено у Таблиці 2.1. Повну таблицю наведено за покликання у [Додатку 1](#).

Словоформа	Лексема	Нормативний відповідник	Транскрипція	Омонімія	Тип суржик у	Аудіозапис	Інформація про мовця	Текст	Примітки
большой	больший	більшій	[б ^о ·о·л ^ш ·о·ї]		лек	1	ж, м. Київ	1	
в общем	в общем	загалом	[в ^о ·обшчем]		лек	1	ж, м. Київ	1	
в общем	в общем	загалом	[в ^о ·обшчем]		лек	1	ж, м. Київ	1	
важен	важний	важливий	[в ^а ·ажен]		лек	1	ж, м. Київ	1	
важно	важно	важливо	[в ^а ·ажн ^о]		лек	1	ж, м. Київ	1	
важно	важно	важливо	[в ^а ·ажн ^о]		лек	1	ж, м. Київ	1	
вигораніє	вигораніє	вигорання	[виг ^о ·ор ^а ·н ^і ·ї·е]		лек	1	ж, м. Київ	1	
вигораніє	вигораніє	вигорання	[виг ^о ·ор ^а ·н ^і ·ї·е]		лек	1	ж, м. Київ	1	
года	год	роки	[г ^о ·ода]		лек	1	ж, м. Київ	1	

Таблиця 2.1. Фрагмент бд корпусу суржикових текстів

Інформація про мовця. Метамова інформація у корпусі слугує для наведення короткої характеристики мовців за статтю та регіоном походження. До неї також входить і вік дикторок, але у самій таблиці він не використовувався, тому що, як було зазначено раніше, одна й та сама респондентка у різні роки мала різний вік.

Аудіозапис, текст. Аудіозаписи та текст у таблиці вказані для кожної суржикової словоформи. До кожної назви прив'язаний відповідний їй аудіо- чи текстовий файл. Це потрібно для того, щоб можна було прослухати вимову суржикової лексики або ж подивитися на контекст, використовуючи текстовий варіант аудіозапису.

Словоформа. Словоформи розташовані в алфавітному порядку у тій формі, у якій вони зустрілися у тексті. Самі тексти та аудіозаписи відсортовані за порядковими номерами їх назв від 1 до 135. До словоформи наведена її лексема, відповідник, транскрипція, розмежування омонімії, тип суржик у та інші примітки. Як результат – отримано словник словоформ, обсягом 2931 одиниця. Його можна використовувати як окремо для обчислення частоти вживання, так і

разом, наприклад, із колонкою «нормативний відповідник» для розробки програми, що автоматично виправлятиме помилки у текстах із суржиком.

Лексема. Для більшості словоформ у корпусі наведено початкову форму, виняток становлять лише певні словосполучення або синтаксичні конструкції, що при зміні втратять свій зміст (наприклад, не при чому – ні до чого, по крайній мері – принаймні). Для іменників, прикметників та числівників початкова форма – це називний відмінок однини (*мальчика – мальчик, особенного – особенний, одною – один*), для прислівників, сполучників, прийменників – та сама форма, що використана у тексті (*отлично – отлично, но – но, с – с*), для дієслів – інфінітив (*работает – работать, пойду – пойти, отношу – относить*). Словник лексем також можна використовувати для подальших досліджень – зокрема, є можливість автоматично розширити парадигму відмінювання та навести всі словоформи для лексеми.

Нормативний відповідник. Для того, щоб підібрати літературний варіант, були використані такі словники української мови:

- одинадцятитомний «Словник української мови» (або СУМ-11). Словник видали у 1970 – 80-х роках, він є першим тлумачним словником української мови. Цей словник використаний за основу у багатьох сучасних українських словниках, наприклад, у «Великому тлумачному словнику сучасної української мови» (ВТССУМ). Після більшості тлумачень у словнику наведено справжні випадки вживання слова, що ілюструють його значення. На сайті опубліковані томи від 1 до 13. [30]
- СЛОВНИК.ua. У словнику міститься понад 130 000 тлумачень із СУМ-11 та понад 21 000 тлумачень, доданих командою та користувачами СЛОВНИК.ua. СЛОВНИК.ua також має Помічника, у якому є база «антисуржика» на понад 700 слів та виразів. А ще є сервіс «Наголоси», що розставляє наголоси в українських текстах [31].

Перевагою для підбору відповідника є також і наявність контексту у текстах. Оскільки для багатьох слів можлива омонімія як у значенні, так і у граматичних формах, контекст спростив задачу наведення відповідників.

Транскрипція. Майже усі суржикові елементи у корпусі вимовлені з використанням засобів як української, так і російської мов. Наприклад, відповідно до фонетичних правил російської мови, більшість приголосних є м'якими (лішня - [л'`ішн'а], некторис - [н'·`ек⁰от⁰орий'е], вместе - [вм'·`е'с'т'е]). В українській мові звуки у більшості випадків є твердими і це також відбито у деяких транскрипціях суржикових слів. Інший приклад – звуки [ч] і [щ]. Вони є твердими в українській мові і м'якими в російській. Тут теж спостерігається мікс вимови: короче - [к⁰ор⁰оче], пєчально - [п'·е·ч'·`а·л'н⁰о]. Російське правило «акання» призводить до варіативності вимови: короче - [к⁰ор⁰оче] і кароче - [кар⁰оче]. Звуки [и] і [е] стають м'якими у російській мові. Використання цього варіанту вимови є найпоширенішим у суржикових словах (саме тому і приголосні стають м'якими). Але це правило зберігається не у всіх випадках: ідеальна - [ід'·е'а·л'на], парєнь - [п'а·р'·е·н'], але привикла - [при^ев`икла], заманчиво - [зам`анчив⁰о].

У транскрипціях використані символи для позначення різних явищ:

- ` - наголос, ставиться перед голосним звуком, на який падає наголос у слові (хуже - [х⁰уже]);
- ' – позначає м'який/пом'якшений приголосний у словах; часто використовується, тому що багато голосних звуків, які пом'якшують попередні приголосні (лежать - [леж`а·т']);
- ⁰ – відбиває явище лабіалізації приголосних перед огубленими голосними [о] та [у] (понімаю - [п⁰о·н'ім`а·ж⁰у]);

- · – позначає і-подібну артикуляцію, коли м'який приголосний звук впливає на артикуляцію сусідніх голосних, що робить їх вимову подібною до вимови звука [i] (*делать* - [d'·`ела·т']);
- : - позначає подовження приголосних (*іменно* - [im'·ен⁰:o]);
- ^ – використовується для позначення асиміляції голосних, коли ненаголошені [и], [е] та [о] в позиціях перед наголошеним голосним вищого або нижчого підняття уподібнюються до інших голосних; це фонетичне правило, загалом, не є особливо характерним для суржику, лише для деяких слів (*потому* - [n⁰от⁰o^ум⁰у], *переводов* - [n'·е·р'·е^ив⁰од⁰ов]);

Варто відмітити, що акомодативне явище назалізації (коли носові приголосні [м], [н] назалізують сусідні голосні звуки) у транскрипціях не позначено, але наявне. Інші явища асиміляції, дисиміляції, редукції тощо, відповідно відбиті у транскрипціях певних слів, де це потрібно (це може бути як українськомовний, так і російськомовний варіанти).

Омонімія, тип суржику, примітки. Омонімія буває різна. Для корпусу, загалом, характерні два її види – омонімія за значенням та омонімія за граматичними формами. Для першого виду все простіше, адже є тексти, які містять контекст для певної словоформи і завдяки цьому можна визначити значення слова. Це спрощує етап підбору відповідника до словоформи. Омонімія за граматичними формами, загалом, включає у себе омонімію відмінкових закінчень. Часто у різних відмінках іменників збігаються відмінкові закінчення. У цьому випадку контекст також допомагає, тому що у тексті видно, де використовується той чи інший відмінок.

Як уже згадувалося раніше, є три типи суржику – фонетичний, лексичний та синтаксичний. Для кожної словоформи визначено те, до якого типу суржику вона належить. Примітки, загалом, використовувалися для зазначення того, що слово

є сленговим, іменем, власною назвою, аббревіатурою тощо. Таких слів небагато, але вони є. Зокрема, примітка сленг використана для 19 словоформ, примітка аббревіатура для 2, ім'я – для 11, назва країни – для 1, лайка – для 6, місто – для 8.

2.2. Статистична параметризація суржикового мовлення за матеріалами корпусу

Для того, щоб унаочнити отримані результати, було підраховано статистику використання суржику у текстах. Для проведення підрахунків було створено окрему таблицю, у якій містяться колонки:

- порядковий номер лексеми;
- лексема;
- абсолютна частота (лексема);
- тип суржику;
- абсолютна частота (тип суржику);
- аудіозапис;
- інформація про мовця;
- текст;
- кількість слів у тексті.

Зазначимо, що колонки *Лексема*, *Тип суржику*, *Аудіозапис*, *Інформація про мовця* та *Текст* були отримані із самого Корпусу суржикових текстів. У Таблиці 2.2 представлено фрагмент бази даних для підрахунків.

Порядковий номер лексеми	Лексема	Абсолютна частота	Тип суржику	Абсолютна частота	Аудіозапис	Інформація про мовця	Текст	Кількість слів у тексті
1	абстрагіруватися	1	лек	2603	83	ж, Київ. обл.	83	89
2	активний	2	фон	305	42	ж, м. Київ	42	124
2	активний	2	лек	2603	41	ж, м. Київ	41	91

3	активно	1	лек	2603	97	ж, Київ. обл.	97	78
---	---------	---	-----	------	----	---------------	----	----

Таблиця 2.2. Фрагмент бд суржикових текстів для підрахунку статистичних даних

Усю таблицю можна переглянути за покликанням у [Додатку 5](#).

Абсолютна частота лексем. Підрахунок абсолютної частоти лексем проводився вручну на матеріалах Корпусу суржикових текстів. Для кожної лексеми у колонці поряд наведено число, що відбиває її абсолютну частоту вживання в усіх текстах.

Загалом, реєстр лем становить 973 одиниці. Десять найчастотніших лексем є такими: *тіпа* (218), *короче* (74), *то єсть* (65), *потому* (54), *вообще* (49), *понімають* (48), *но* (45), *щас* (42), *потом* (38), *общаться* (36). Найменш частотними є лексеми *любительський* (1), *таракан* (1), *больший* (1), *важний* (1), *ісполняться* (1), *оставлять* (1), *получать* (1), *репетиція* (1), *учітуя* (1), *часть* (1) тощо.

Співвідношення абсолютних частот та кількості лексем можна побачити у Таблиці 2.3.

Абсолютна частота	Кількість лексем
1	569
2	181
3	77
4	27
5	21
6	12
7	16
8	9
9	7
10	3
11	1

12	3
13	2
14	2
15	2
17	1
20	4
21	1
26	2
29	1
31	1
32	1
35	1
36	1
38	1
42	1
45	1
48	1
49	1
54	1
65	1
74	1
218	1

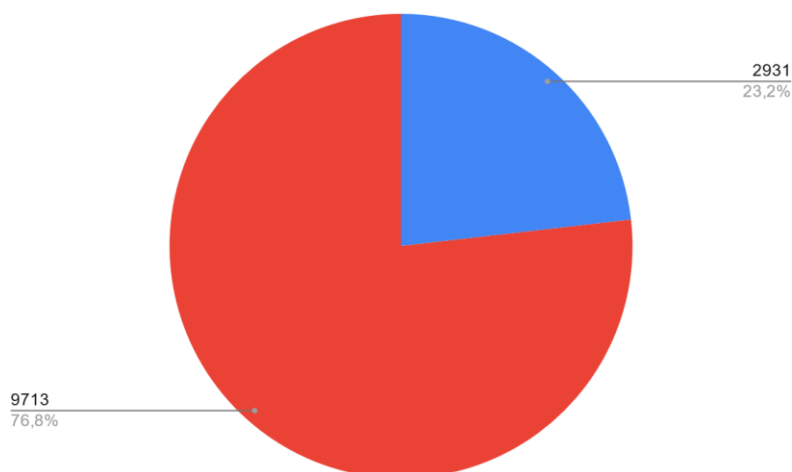
Таблиця 2.3. Співвідношення абсолютних частот та кількості лексем

Аналізуючи створену таблицю, можна зробити висновок, що найбільше лексем (569) мають частоту 1. Зазвичай, це правдиво для текстів або звичайного мовлення – чим більше низькочастотних лексем, тим краще і повніше виражена думка автора. Але для дослідження суржику цей аспект можна інтерпретувати по-іншому: серед 973 лексем є 569 таких, що є унікально витвореними мовцями у процесі монологу. Щодо найчастотніших – це або вставні слова (*тіна, короче,*

то єсть), або сполучники (но, тому), або дієслова (понімають, об'язуються). Частотне використання дієслів зумовлене, насамперед, контекстом.

Відносна частота лексем. Для підрахунку відносної частоти використання суржикової лексики у всіх 135 текстах було використано формулу $h = \frac{t}{N} \times 100\%$, де h – відносна частота, t – абсолютна частота суржикових слововживань у текстах, N – загальний обсяг слововживань у всіх текстах.

Отже, $N = 12\,644$ слововживання для 135 текстів, $t = 2931$ (сума абсолютних частот для 943 лексем), тому $h = \frac{2931}{12\,644} \times 100\% = 23,18\%$. Відповідно, розподіл нормативної та суржикової лексики у текстах відображено на Діаграмі 2.1.

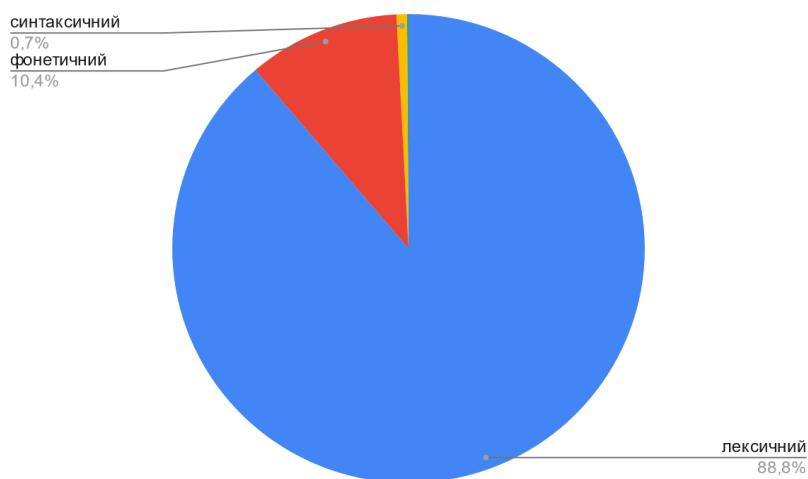


Діаграма 2.1. Відсоткове співвідношення нормативної та суржикової лексики.

З Діаграми 1 видно, що суржикова лексика становить майже четвертину від усіх слововживань у текстах. Майже кожне четверте слово у монологіях дикторів є суржиковим, що свідчить про те, що використання суржика як мовного засобу є характерним для них. Важливо відмітити, що з іншого боку, цей результат не є настільки поганим; 9713 із 12 644 словоформ були нормативними, а тому рівень

мовних знань для використання тільки нормативної лексики є доволі високим.

Відносна частота типів суржику. Для її обчислення підраховано кількість елементів, які мають ознаки лексичного, фонетичного та синтаксичного типів суржику. Загальна кількість суржикових слів становить 2931, серед неї 2603 елементи належать до лексичного типу, 305 – до фонетичного, 21 – до синтаксичного і 3 – до інших типів (поєднання рівнів; фонетичний + синтаксичний або лексичний + синтаксичний). Співвідношення кількості відображено на Діаграмі 2.



Діаграма 2.2. Співвідношення різних типів суржику у корпусі

Більша частина слів у корпусі відносяться до лексичного типу суржику, оскільки вони є запозиченнями із російської мови, менше чи більше викривленими під українську вимову. Наприклад, *вообще* - [в⁰о:бщ'е], *окончательно* - [ок⁰онч`а-т`-е-л`н⁰о], *интересуюсь* - [інт`-е-р`-ес⁰у-ж⁰у-с']. Менша частина є фонетичним типом, зазвичай, це просто пом'якшена вимова звука, який повинен бути твердим, або вживання флексії –ть замість –ти для інфінітиву: *реально* - [р`-е`а-л`н⁰о], *спать* - [сп`а-т']. Слів і словосполучень синтаксичного типу ще менше, оскільки у корпус виносилися переважно тільки

словоформи. До цього типу можна віднести неправильне відмінкове закінчення або сталий вираз, характерний для російської мови, але не для української: *в Тік токе* - [вт'ікт^ооке], *до сіх пор* - [д^оос'іхп^{ор}].

Висновок до розділу II

Суржик є мовним гібридом, тому працювати з ним доволі важко. Для проведення дослідження було відібрано 135 аудіозаписів, загальна тривалість яких становить 6041 секунду. Після цього вони були відповідно опрацьовані для створення текстгйдів та файлів із текстовим поданням записів. При цьому виникли деякі проблеми. Розпізнавання слів у аудіозаписах є складним, тому що інколи слова зашумлені, їх погано чути, неможливо встановити правильну форму або варіант написання. Можливо сприймати ситуації з деякими словами по-різному. Наприклад, візьмемо слово *пепел*. З одного боку, це просто лексична калька російського слова *пепел*. Але можлива й інша ситуація – мовець намагається передати російську вимову цього слова засобами української мови. На виході отримуємо приблизно однакове звучання. Це викликає певні труднощі при розпізнаванні мовлення, а також у інтерпретації того, до якого рівня можна віднести ті чи інші слова.

Після опрацювання записів, було створено базу даних для суржикізмів. У ній міститься 2931 суржикова словоформа та, відповідно, 943 лексеми. До всіх словоформ наведено нормативний відповідник, транскрипцію, тип суржику, зняття омонімії (де потрібно), а також примітки. Варто зазначити, що кожен суржикізм прив'язано до відповідного аудіозапису та тексту, з якого його отримано, для того, щоб можна було послухати вимову та переглянути контекст за потреби.

Статистично, суржикові словоформи становлять 23,2% від усіх слововживань у корпусі. Це, фактично, означає, що майже четвертина від усіх слововживань є викривленою російсько-українською лексикою. Але не можна сказати, що це

багато, тому що інша частина – нормативна лексика, якою вільно володіють усі дикторки.

Із написанням транскрипцій також пов'язані певні проблеми. Інколи слово настільки перекручене, що передати його звучання дуже складно. Загалом, усі транскрипції розписані із урахуванням особливостей вимови за диктором та орфоепічних норм української та російської мов.

При підборі нормативного відповідника варто враховувати контекст та саму форму суржикізму. Не завжди можливо встановити точний відповідник, що передав би значення слова і замінив би його. Інколи для такого потрібно переробити ціле речення, тому що, наприклад, рід іменника не співпадає. Ще одне, про що варто згадати, – відсутність деяких форм у російській або українській мовах. Іноді в російській мові є одне конкретне слово для позначення якогось поняття, а в українській мові це поняття можливо передати лише описово. Наприклад, сур. *оживленіє*, рос. *оживление*, укр. *робити когось живим*.

Лексичні суржицизми – найбільш поширені в корпусі. Вони займають 88,8% від усіх типів суржику. Також можна сказати, що саме лексичний тип постійно поповнюється новими варіаціями, адже суржик, як і будь-яка мова, розвивається. Лексика найбільше трансформується теж на цьому рівні – слова скорочують та перекручують різними способами.

Варто відмітити, що сам корпус може бути і далі розширений шляхом підбору нових аудіозаписів та їх опрацюванням. Також можливо використовувати його і для вирішення інших лінгвістичних завдань – наприклад, розробки програми для автоматичного виправлення суржику у зібраних текстах.

ВИСНОВКИ

Проведене дослідження присвячено особливостям суржику як мовного явища, що постало на основі української та російської мов. Центральним поняттям дослідження є поняття суржику, що витворюється за певних умов та протиставляється нормативним мовам: українській та російській. Робочим визначенням суржику у роботі є таке визначення: «Суржик – мовне явище, якому характерне ненормоване змішування фонетичних, лексичних та граматичних особливостей російської та української мов».

Суржик як явище усного мовлення з'явився у кінці XVII століття серед чиновництва та солдатства. Остаточно він закріпився у радянські роки, коли заборонялося використовувати українську мову. Проводилась заміна власне українських слів на слова, близькі до російської мови. Фактично це призвело до того, що суржик закріпився навіть у письмовій формі у різних книжках.

Основними причинами використання людиною суржику є бажання бути зрозумілим для іншої людини, а також бажання бути «своїм». Мовець підлаштовується під розмову, зовсім не задумуючись про чистоту своєї мови. Існують й інші причини – двомовність у сім'ї, школі та серед однолітків; незнання меж у використанні засобів двох мов тощо.

Базуючись на найбільш поширених різновидах суржику, можливо виділити декілька його типів: фонетичний, лексичний та синтаксичний. Фонетичний тип суржику виникає при засвоєнні чужих фонетичних законів та витворенні спотворених по звучанню слів. Лексичний тип – найпоширеніший; цей тип характеризується запозиченням лексики та значень слів, а також лексичним калькуванням, що дає простір для фантазії. Синтаксичний – найбільш складний для відображення, адже він виникає лише у словосполученнях та реченнях, тобто за наявності певного контексту; граматичні правила однієї мови переносяться на іншу.

Для створення корпусу та проведення дослідження суржикової лексики, було опрацьовано 135 аудіозаписів формату .wav різної тривалості – від 15 секунд до 5 хвилин. Загальна тривалість усіх оброблених аудіозаписів становить годину сорок хвилин та сорок секунд, тобто, приблизно 6041 секунду усного мовлення. Для кожного аудіозапису подано текстгрід – спеціально створений у програмі Praat файл, який містить анотацію до аудіозапису із різними примітками про вимову, хезитації тощо. Найважливішою позначкою у текстгрідах є позначка *c* (суржикове мовлення), оскільки саме завдяки їй пізніше укладався реєстр суржикізмів. До кожного аудіозапису також подано файли у форматі .docx, що містять текстове представлення сказаного дикторами. Загальний обсяг текстів (тобто, і самого корпусу) становить 12644 слововживання. Після цього, власне, почалась робота із укладання самого корпусу. Реєстр суржицизмів формувался на основі отриманих файлів опрацьованих аудіозаписів. Суржикові словоформи вносилися у таблицю у такій формі, як були вжиті самими мовцями. Їх кількість дорівнює 2931 словоформа. До всіх словоформ наведено лексему (943 елементи), нормативний відповідник, транскрипцію, тип суржику, зняття омонімії (де потрібно), а також примітки. Варто зазначити, що кожен суржикізм прив'язано до відповідного аудіозапису та тексту, з якого його отримано, для того, щоб можна було послухати вимову та переглянути контекст за потреби.

Створення корпусу дозволило провести і деякі статистичні обрахунки. Зокрема, визначити абсолютну та відносну частоти використаних у тексті лексем, а також відстежити частку входження різних типів суржику у реєстр. Найчастотнішими лексемами є *тіна* (218), *короче* (74), *то єсть* (65), *потому* (54), *вообще* (49), *понімають* (48), *но* (45), *щас* (42), *потом* (38), *общаться* (36). Найменш частотними є лексеми *любительський* (1), *таракан* (1), *больший* (1), *важний* (1), *ісполняються* (1), *оставлять* (1), *получать* (1), *респіція* (1), *учітуя* (1), *часть* (1) *тощо*. Зазначимо, що лексем із частотою 1 – найбільше у корпусі, 569 штук. Це зумовлено особливостями мови, яка тяжіє до унікальності, а також свідчить про

те, що більша частина лексем у корпусі є індивідуально витвореними. Суржикові лексеми займають 23,2% від усіх слововживань у корпусі. Кожне четверте слово у мовленні дикторів є суржикізмом, але всі інші є нормативними українськими словами, що свідчить про достатній рівень знань. Щодо типів суржику – очевидно, найбільшу частку становить саме лексичний тип (88,8 %), оскільки, в основному, лексика у суржику є запозиченою із російської та викривленою під українську вимову. Фонетичний тип проявлявся у неправильній вимові звука, наприклад, пом'якшенні не там, де треба. Синтаксичний тип займає усього 0,7 % від загальної кількості. До нього відносяться якісь сталі вирази, що характерні для російської мови, але не для української.

Корпус суржикових текстів надалі можна використовувати для різних лінгвістичних досліджень. У корпусі містяться два словники – лексем та словоформ. Словник лексем пізніше можна автоматично розширити, вивівши усі словоформи лексеми. Словник словоформ підходить для дослідження частоти та вживаності певних форм слова. Анотовані записи та транскрипції словоформ можна використати для навчання систем розпізнавання мовлення, тому що суржик погано розпізнається та знижує відсоток успішності розпізнавання. Нормативний варіант, підібраний до словоформи, можна використати для автоматичного виправлення помилок у текстах. Також уже створений корпус, за бажанням, можна доповнити новими записами та словоформами.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Антисуржик. Вчимося ввічливо поводитись і правильно говорити: [навч. посіб.] / за заг. ред О. Сербенської; Львів. нац. ун-т ім. І. Франка. – Львів: Світ, 1994. – 152 с.
2. Бацевич Ф. С. Словник термінів міжкультурної комунікації / Бацевич Ф. С. – К. : Довіра, 2007. – 205 с
3. Белей, Лесь Суржик: соціолінгвістична перспектива / Лесь Белей // Сучасні проблеми мовознавства та літературознавства : зб. наук. праць / відп. ред. І.В. Сабадош. – Ужгород, 2017. – Вип.22. – С. 5–8. – Рез. англ. – Бібліогр. : с. 7–8 (17 назв).
4. Бобрусь, М. Є., & Наволокін, І. П. (2023, January). ВИКОРИСТАННЯ СУРЖИКІВ В УКРАЇНСЬКІЙ ПІСЕННІЙ КУЛЬТУРІ. In *The 2th International scientific and practical conference "Modern education using the latest technologies" (January 17-20, 2023) Lisbon, Portugal. International Science Group. 2023. 504 p.* (p. 49).
5. Брага І. І. Українсько-російський суржик в соціо-комунікативній ситуації ринку. Мова і суспільство: зб. ст. / ред. кол.: Г. Мацюк. Львів: ЛНУ ім. І. Франка, 2010. Вип. 2. 2011.
6. Бриних М. "Шахмати для дибілів" [Електронний ресурс] / М. Бриних. – 2008. – Режим доступу до ресурсу: https://chtyvo.org.ua/authors/Brynykh_Mykhailo/Shakhmaty_dlia_dybiliv_vyd_2008/.
7. Вишня О. "Вишневі усмішки сільські" [Електронний ресурс] / Остап Вишня – Режим доступу до ресурсу: <https://www.ukrlib.com.ua/books/printit.php?tid=3027>.
8. Вступ до мовознавства: підручник/ І. О. Голубовська, С. М. Лучканин, В. Ф. Чемес та ін.; за ред. І. О. Голубовської. Київ: ВЦ «Академія», 2016. 320с
9. Герман В. В. СУРЖИК ЯК НАСЛІДОК ІНТЕРФЕРЕНЦІЇ / В. В. Герман // Українознавчий вимір у сучасній науці: гуманітарний аспект: матеріали ІV

- Всеукраїнської науково-практичної конференції, 13 квітня 2018 р., м. Миколаїв. / В. В. Герман. – Миколаїв, 2018. – С. 39 – 42.
10. Голощук, С. Л. (2022). КОРПУСНА ЛІНГВІСТИКА: СУЧАСНИЙ СТАН ТА ПЕРСПЕКТИВИ ДОСЛІДЖЕНЬ CORPUS LINGUISTICS: MODERN APPROACH AND RESEARCH PERSPECTIVE. *ФІЛОЛОГІЧНІ СТУДІЇ*, 249.
 11. Гофендер Філіп. Радянська мовна політика в Україні. Переклади творів Леніна українською 1930-х та 1950-х рр.// Дивослово. – 2010. – Ч. 4. – С. 37 – 47.
 12. ГРИМАШЕВИЧ, Г. І. Новий український правопис: історія, сучасність, перспективи. *Формування базових і фахових компетентностей студентів у закладах вищої освіти: Матеріали науково-практичної конференції*, 2020, с. 26-28.
 13. Грицак Я. Нариси історії України: формування модерної української нації XIX-XX ст. [Навч. посібник]. - Київ: Генеза, 1996. - 360 с.
 14. Грінченко Б. "Серед темної ночі" [Електронний ресурс] / Б. Грінченко – Режим доступу до ресурсу: <https://www.ukrlib.com.ua/books/printit.php?tid=803>.
 15. Дзюбишина-Мельник Н. Я. Суржик і суржикізми : стилістичні ресурси. Наукові записки НаУКМА. 2010. Т. 111 : Філологічні науки. С. 16–20.
 16. Жуковська В. В. Вступ до корпусної лінгвістики: навчальний посібник / В. В. Жуковська – Житомир: Вид-во ЖДУ ім. І. Франка, 2013. – 142 с.
 17. ЗАВОДНА, Леся. Суржик: мовна патологія чи норма українського суспільства?. *New pedagogical thought*, 2020, 104.4: 50-53.
 18. Капінус, О. Л. (2022). ВИКОРИСТАННЯ ТЕКСТОВИХ КОРПУСІВ У ЛІНГВІСТИЧНИХ ДОСЛІДЖЕННЯХ THE USE OF THE TEXT CORPORA IN LINGUISTIC RESEARCH. *ФІЛОЛОГІЧНІ СТУДІЇ*, 234.

19. Левченко Тетяна. 2019. Українські мовознавці про суржик у мові засобів масової комунікації. „Науковий вісник Міжнародного гуманітарного університету / Серія: Філологія 40” т. 3: 106–109
20. Макарець Ю.С. Державна мовна політика незалежної України. Науковий вісник Міжнародного гуманітарного університету. Серія «Філологія». 2019. № 40. Том 1. 53–58.
21. Масенко Л. Т. Суржик : між мовою і язиком / Лариса Масенко. - 2-ге вид., зі змінами і допов. - Київ : Києво-Могилянська академія, 2019. - 229 с. : портр.
22. Мельничайко, В., & Криській, М. (2017). Порухення лексичних норм – найхарактерніша ознака суржику.
23. Мірошниченко, П. В. (2017). Суржик у звуковому образі українського радіомовлення.
24. Міхаель Мозер. ”Суржик” чи ”суржики”? / Міхаель Мозер // Українська мова. — 2016. — № 1. — С. 27-54.
25. Про забезпечення функціонування української мови як державної - Закон України від 25.04.2019 № 2704-VIII [Електронний ресурс]. – 2019. – Режим доступу до ресурсу: <https://zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=2704-19>
26. Проблема форм міксації мов у соціолінгвістиці / О. Селіванова // Мова і суспільство. - 2011. - Вип. 2. - С. 3–10. - Режим доступу: http://nbuv.gov.ua/UJRN/mis_2011_2_3
27. Радчук В. Суржик як недопереклад/ / Укр. мова та літ.— 2000.— № 11 (171)
28. С. Дель Гаудіо, Взаємодія між суржигом і діалектами [в:] „Мова і культура”, nr 13/8(144), 2010
29. Словник української мови: [в 11 т.] / АН Української РСР, Ін-т мовознав. ім. О. О. Потебні ; редкол.: І. К. Білодід (голова) [та ін.]. - Київ : Наук. думка, 1970 - 1980.

- 30.Словник української мови online. URL: <https://sum20ua.com>
31. Словник.ua. URL: <https://slovnyk.ua/>
- 32.Тараненко, К. В. СУРЖИК ЯК ЕЛЕМЕНТ ВЕРБАЛЬНОГО ІМІДЖУ УКРАЇНСЬКОГО ВИКОНАВЦЯ. *ВЧЕНІ ЗАПИСКИ*, 1202239.
- 33.Типи інтерферентних впливів у білінгвальному середовищі / В. М. Труб // Наукові записки НаУКМА. Філологічні науки. - 2012. - Т. 137. - С. 142-147. - Режим доступу: http://nbuv.gov.ua/UJRN/NaUKMAfn_2012_137_36
- 34.Томіленко, Л.М. Суржик як об'єкт наукових досліджень і дискусій / Л.М. Томіленко // Мовознавство. - 2014. - № 4. - С. 69-81.
- 35.Узунов С. Мовна ситуація в Україні: білінгвізм чи диглосія? *Українське мовознавство*. Київ, 2014. Вип. 1 (44). С. 81 – 93.
- 36.Українська мова: енциклопедія / [редкол. : Русанівський В.М.(співголова), Тараненко О.О.(співголова), Зяблюк М.П.та ін.]. –2-ге вид., випр. і доп. – К. : Вид-во “Українська енциклопедія” ім. М.П.Бажана, 2004.
- 37.Чех А. Цього ви не знайдете в Яндексі [Електронний ресурс] / А. Чех. – 2006. – Режим доступу до ресурсу: https://chtyvo.org.ua/authors/Cekh_Artem/Tsoho_vy_ne_znaidete_v_Yandex_si/.
- 38.Шевченко Т. Г. Комедія "Сон" [Електронний ресурс] / Т. Г. Шевченко – Режим доступу до ресурсу: <http://litopys.org.ua/shevchenko/shev128.htm>.
- 39.Шевчук-Клюжева, О. В. (2016). Суржик в Інтернет-просторі (на матеріалі соціальних мереж Вконтакте, Facebook, Instagram).
- 40.Adolphs, S., & Knight, D. (2010). Building a spoken corpus. *The Routledge handbook of corpus linguistics*, 38-52
- 41.BILANIUK, L. (1997). Speaking of "Surzhyk": Ideologies and Mixed Languages. *Harvard Ukrainian Studies*, 21(1/2), 93-117. Retrieved April 8, 2021, from <http://www.jstor.org/stable/41036643>
- 42.Flier M. Surzhyk: The Rules of Engagement.// *Cultures and Nationalisms of East and Cen-tral Europe. Essays in Honour of Roman Szporliuk.*/ Ed. By Zvi

- Gitelman et al. [= Harvard Ukrainian Studies Vol. XXII, 1998].
Cambridge/Mass., 2000. — C. 113–129.
43. Hentschel, G., & Palinska, O. (2022). Restructuring in a mesolect: a case study on the basis of the formal variation of the infinitive in Ukrainian-Russian surzhyk. *Cognitive Studies, Études cognitives*, 22.
44. Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.
45. Kent, Kateryna. (2012). *Morphosyntactic analysis of Surzhyk, a Russian–Ukrainian mixed lect.*. Retrieved from the University of Minnesota Digital Conservancy, <https://hdl.handle.net/11299/137718>.
46. McEnery, T. (2019). *Corpus linguistics*. Edinburgh University Press.

ДОДАТКИ

1. Додаток 1. Корпус суржикових текстів. Режим доступу:
https://docs.google.com/spreadsheets/d/1wPLsC2GTcYwBtp-YVIPg7Z1jTgxJGtZW1_kK-VEghXY/edit?usp=sharing
2. Додаток 2. Аудіозаписи. Режим доступу:
https://drive.google.com/drive/folders/1TMM4WYdOVKt9NEFZxyoFog0oBAt0diww?usp=share_link
3. Додаток 3. TextGrid. Режим доступу:
https://drive.google.com/drive/folders/1x3PBumMRjaZPFJAoKvceRSdMJ6dlynj6?usp=share_link
4. Додаток 4. Тексти. Режим доступу:
https://drive.google.com/drive/folders/11ToJWDWEh55Jv4d2Kp48aGZT5_NuunQs?usp=share_link
5. Додаток 5. Корпус суржикових текстів: для статистики. Режим доступу:
https://docs.google.com/spreadsheets/d/1AOBzIhBupp-1BiszvQxa6xgFmb4ydrNM9klt-AtjXE/edit?usp=drive_link