

Міністерство освіти і науки України
Національний університет «Києво-Могилянська академія»
Факультет інформатики
Кафедра математики

Магістерська робота

освітній ступінь – магістр

на тему: **«Математичне моделювання та прогнозування зміни кількості
носіїв української мови та інших мов в Україні»**

Виконала: студентка 2-го року
навчання

освітньо-наукової програми
«Прикладна математика»,
спеціальності 113 Прикладна
математика

Будишевська Марина
Олександрівна

Керівник: Дутка В. А.,
кандидат техн. наук, доцент

Рецензент _____
(прізвище та ініціали)

Магістерська робота захищена
з оцінкою _____

Секретар ЕК _____

Київ – 2021

Тема: Математичне моделювання та прогнозування зміни кількості носіїв української мови та інших мов в Україні.

Календарний план виконання роботи:

№ п/п	Назва етапу курсової роботи	Термін виконання етапу	Примітка
1.	Отримання теми курсової роботи.	09.10.2020	
2.	Пошук та збір тематичних матеріалів.	25.11.2020	
3.	Ознайомлення з матеріалами та обдумування структури роботи.	30.12.2020	
4.	Написання вступу та плану роботи.	18.01.2021	
5.	Вивчення матеріалів, які необхідні для написання першого розділу роботи, написання першого розділу .	12.02.2021	
5.	Опрацювання матеріалів та написання другого розділу роботи.	21.02.2021	
6.	Вивчення джерел та завершення роботи над другим розділом роботи.	03.03.2021	
7.	Правильне оформлення роботи відповідно до вимог написання курсової роботи.	04.03.2021	
8.	Створення презентації та написання доповіді для захисту роботи.	10.04.2021	
9.	Подання та аналіз попередньої версії роботи з керівником.	21.04.2021	
10.	Коригування роботи згідно із зауваженнями керівника.	21.05.2021	
11.	Захист курсової роботи.	10.06.2021	

Студентка Будишевська М. О.

Керівник Дутка В. А.

“ ”

ЗМІСТ

Анотація	4
Вступ	5
РОЗДІЛ 1 ОГЛЯД ДЕЯКИХ ВІДОМИХ МОДЕЛЕЙ	7
1.1. Модель Лотки-Вольтерри	7
1.2. Аналог реакційно-дифузійної системи	8
1.3. Висновки за розділом 1	11
РОЗДІЛ 2 ЗАСТОСУВАННЯ МОДЕЛЕЙ ДО ПРОЦЕСУ ЗМІНИ КІЛЬКОСТІ НОСІЇВ МОВИ В УКРАЇНІ	12
2.1. Застосування системи Лотки-Вольтерри	12
2.2. Застосування трикомпонентної нелінійної математичної моделі. Підготовка даних	14
2.3 Програмна реалізація аналогу реакційно-дифузійної системи за допомогою Python	17
Висновки	23
Список використаних джерел	24
Додатки	26

Анотація

У роботі розглянуто дві математичні моделі та використано саме трикомпонентну нелінійну математичну модель для прогнозування зміни кількості носіїв мови в Україні, що описує взаємодію спільноти, у якій індивіди розмовляють декількома мовами та деколи переходять із рідної мови на користь іншої та навпаки – яка є більш відповідною для отримання точних результатів.

Ключові слова: математичне моделювання, реакційно-дифузійна система.

ВСТУП

Математичне моделювання – процес, у якому за допомогою математики можна описати взаємозв'язок у реальних життєвих ситуаціях та, власне, самі ситуації. Такі моделі виходять за межі фізичних характеристик ситуації для вивчення її структурних особливостей за допомогою математики, що тягне за собою побудову моделей природних і соціальних явищ, і вибір релевантної математики є частиною процесу вирішення проблеми [12].

Математику не слід розглядати як науку лише для науки, оскільки в наш час математично-статистичний аналіз може широко застосовуватися в соціально-політичній сфері, виборчих кампаніях та різних аспектах життя.

Магістерська робота зосереджена на забезпеченні цілісного розуміння тенденцій та закономірностей використання української мови. Такі знання можуть забезпечити довгостроковий демократичний розвиток суспільства, оскільки мовне питання є питанням ідентичності та є одним із ключових питань у процесі національного будівництва, самовизначення та розвитку держави.

Протягом останніх років дане питання набуло критичного значення через військовий конфлікт у країні. Це було політизовано і використано для виправдання військової агресії проти нашої країни. Тому важливість мовного питання визначається не лише культурою, але й політичними потребами та проблемами. Крім того важливо відстежувати та виявляти мовні зміни та динаміку, щоб проаналізувати особливості формування сучасної політичної нації. Розуміння таких особливостей може бути використано у постійному демократичному розвитку держави.

Динаміку мовних змін можна описати та спрогнозувати за допомогою дифузійних систем виведених А.Кендлером, Р.Юнгером та Дж.Стілем [10].

Мета роботи. Дослідити і проаналізувати трикомпонентну реакційно-дифузійну систему моделювання конкуренції та за допомогою Python спрогнозувати кількість носіїв мови в Україні.

Об'єкт дослідження. Математична модель, увага якої спрямована на визначення тенденції зміни кількості носіїв мови.

Робота складається зі вступу та двох розділів.

Перший розділ складається із загальних відомостей про модель Лотки-Вольтерри, аналог реакційно-дифузійної системи.

У другому розділі описано практичне застосування моделі Лотки-Вольтерри; за допомогою нелінійної трикомпонентної моделі показано процес підготовки даних для подальшого прогнозування зміни кількості носіїв мови в Україні, етапи реалізації програмного коду та візуалізація отриманих результатів із висновками.

Робота може бути корисною для відстеження та виявлення мовних змін в країні. Розуміння таких особливостей є важливим у контексті демократичного розвитку, побудови державної політики та відслідковування соціальних процесів.

РОЗДІЛ 1 ОГЛЯД ДЕЯКИХ ВІДОМИХ МОДЕЛЕЙ

Під зміною кількості носіїв мови розуміють такий процес, коли частина спільноти, у якій розмовляють декількома мовами, відмовляється від рідної мови на користь іншої та навпаки. У даному випадку, розглядатимемо динаміку зміни кількості носіїв мови як конкурентний процес, коли кількість носіїв мови постійно змінюється (як результат смертності або народжуваності, імміграції або еміграції) [10].

Спробуємо підібрати декілька моделей, які за тих чи інших припущень будуть описувати процес зміни кількості носіїв мови.

1.1 Модель Лотки-Вольтерри

Модель Лотки-Вольтерри (або рівняння хижак-жертва) — модель, яка описується двома диференціальними рівняннями, і описує кінетику чисельності популяції із одним типом хижаків і одним типом жертв. Рівняння запропонували незалежно Альфред Джеймс Лотка та Віто Вольтерра, в 1925 та 1926 роках, відповідно [6].

Розглянемо базову модель Лотки-Вольтерри, яка має загальний вигляд:

1) зміна популяції жертв описується так:

$$\frac{dN_1}{dt} = r_1 N_1 - p_1 N_1 N_2 ;$$

2) приріст популяції хижаків:

$$\frac{dN_2}{dt} = -d_1 N_2 + p_2 N_1 N_2 ;$$

де N_1 — чисельність жертви у момент часу t ; N_2 — чисельність хижака у момент часу t ; r_1 — коефіцієнт приросту жертви за відсутності хижака(народжуваність); p_1 — коефіцієнт винищування хижаком жертв; d_1 — коефіцієнт смертності хижака

($-d_1$ – коефіцієнт приросту хижака за відсутності жертви); p_2 – коефіцієнт хижацтва для хижака [4,11].

У основі системи присутні наступні гіпотези:

1) за відсутності хижаків жертви розмножуються необмежено

$$N'_1 = r_1 N_1;$$

2) хижаки за відсутності жертв вимирають

$$(N'_2 = -d_1 N_2);$$

3) ефект впливу популяції хижаків на популяцію жертв полягає в зменшенні відносної швидкості приросту чисельності жертв(r_1) на величину, пропорційну чисельності хижаків [4,11].

1.2 Аналог реакційно-дифузійної системи

Розглянемо трикомпонентну нелінійну математичну модель, запропоновану А. Кандлером (див. [10]) яка описує взаємодію спільноти, у якій індивіди розмовляють декількома мовами та деколи переходять із рідної мови на користь іншої та навпаки. Дана модель складається із трьох нелінійних рівнянь реакції-дифузії, які в одновимірному наближенні має наступний вигляд і описує взаємодію трьох спільнот носіїв мови [9,10]:

$$\begin{aligned} u_t &= \lambda_1 u_{xx} + a_1 u \left(1 - \frac{u}{K - (v + w)} \right) - c_{31} u w + c_{12} u v, \\ v_t &= \lambda_2 v_{xx} + a_2 v \left(1 - \frac{v}{K - (u + w)} \right) + (c_{13} + c_{31}) u w - (c_{12} u + c_{32} w) v, \\ w_t &= \lambda_3 w_{xx} + a_3 w \left(1 - \frac{w}{K - (u + v)} \right) - c_{13} u w + c_{32} v w. \end{aligned} \quad (1)$$

Функції $u = u(t, x)$ та $w = w(t, x)$ описують частоту носіїв мови (в нашому випадку $u(t, x)$ відповідає за україномовних носіїв та $w(t, x)$ – російськомовних носіїв, які говорять завжди (або майже завжди) рідною мовою; $v(t, x)$ відповідає за спільноту носіїв, які вільно розмовляють обома мовами(українською та російською) і використовують кожен мову залежно від

обставин(білінгви). Залежні від часу та простору змінні u та v відповідають за частоти одномовних носіїв української та російської мов, у той час як w описує частоту тих самих білінгвів. Похідні часу u_t , v_t і w_t вказують на швидкість зміни цих частот, тоді як похідні простору описують дифузію в просторі мовців. Терміни у правій частині системи рівнянь (1) описують зміни частоти носіїв мови у кожній з трьох підгруп u , v та w [9]. Функція $x = x(t)$ відповідає за кількість населення в Україні залежно від часу t .

Другі змінні в кожному початково заданому рівнянні (1) - є деяким узагальненням стандартних логістичних термінів, що виникають у багатьох добре відомих моделях, включаючи відоме рівняння Фішера та дифузійну систему Лотки – Вольтерри для видів, що взаємодіють [11]. Константа K визначає верхню межу усіх трьох спільнот носіїв, тобто, передбачається, що $u + v + w < K$ [9].

Коефіцієнти c_{13} та c_{31} відповідають за ймовірність зміни мови, що спричиняє білінгвістичність носіїв на основі диференціальної престижності або привабливості двох конкуруючих мов; $c_{31}uw$ та $c_{13}uw$ – описують білінгвів, які вживають обидві мови за обставин. Можна зазначити, що зміна мови призводить до зростання кількості носіїв у двомовній спільноті (за умови відсутності будь-яких інших впливаючих чинників). З іншого боку, терміни $c_{12}uv$ та $c_{32}vw$ описують протилежну тенденцію, коли білінгва може стати монолінгвою. Це може трапитись, наприклад, у випадку державної політики, яка призведе до зниження статусу однієї мови порівняно з іншою, де реальним прикладом може слугувати Валуєвський циркуляр, у якому наказувалося призупинити видання значної частини книг, написаних «малоросійською», тобто українською мовою [1]. Коефіцієнти c_{12} та c_{32} репрезентують ймовірність того, що білінгви можуть стати монолінгвами у спільноті u та w , відповідно. Варто зазначити, що нерівність $c_{12} < c_{32}$ (зокрема, якщо $c_{12} \ll c_{32}$, тоді $c_{12} = 0$) має місце тоді, коли мова спільноти u знаходиться під тиском [9].

Двомовна субпопуляція рекрутує з обох одномовних підгруп населення зі швидкістю $(c_{13} + c_{31})uw$. У свою чергу, білінгви переходять на одномовність і використовують ту чи іншу мову зі швидкістю $c_{12}u_1u_2$ та $c_{32}u_3u_2$ [10].

Запропонована модель у першоджерелах була використана для моделювання процесу англізації у Шотландії протягом 20 століття, у результаті чого відсоток гальської мови в усіх частинах Шотландії різко зменшився (система була вирішена чисельно, для демонстрування хорошої відповідності між числовими рішеннями та офіційними статистичними даними). Надалі розглянемо модифікацію трикомпонентної системи (1) та властивості отриманих розв'язків, звівши у системі (1) дробові нелінійності до квадратичних (дробові нелінійності, що виникають у (1), є прямим узагальненням тих, що були введені в роботі [10]). Передбачається, що носії обох мов мають спільну несучу здатність K . Це припущення до кінця не є обґрунтованим, оскільки мова певного мовця, як правило, пов'язана з національністю, наприклад. Отже, не можна стверджувати, що різні національності мають однакову несучу здатність весь час [9].

Основна модель [10] містить стандартні логістичні терміни, що виникають у багатьох біологічних аспектах мотивованих моделей [9]. Замінімо дробові нелінійності за логістичними термінами, які також обмежують необмежений ріст цих спільнот. Це значить, що $\frac{u}{K-(v+w)}$, $\frac{v}{K-(u+w)}$ та $\frac{w}{K-(u+v)}$ замінюються на $\frac{u}{K_1}$, $\frac{v}{K_2}$ та $\frac{w}{K_3}$ відповідно. Маємо систему, що містить лише квадратичні нелінійності [9]:

$$\begin{aligned} u_t &= \lambda_1 u_{xx} + a_1 u \left(1 - \frac{u}{K_1}\right) - c_{31}uw + c_{12}uv, \\ v_t &= \lambda_2 v_{xx} + a_2 v \left(1 - \frac{v}{K_2}\right) + (c_{13} + c_{31})uw - (c_{12}u + c_{32}w)v, \\ w_t &= \lambda_3 w_{xx} + a_3 w \left(1 - \frac{w}{K_3}\right) - c_{13}uw + c_{32}vw. \end{aligned} \quad (2)$$

Нехай коефіцієнти λ_i , a_i та K_i ($i = 1, 2, 3$) мають позитивне значення (усі інші невід'ємні і деякі із них можуть дорівнювати нулю).

Отриману систему можна спростити:

$$u \rightarrow K_1 u, v \rightarrow K_2 v, w \rightarrow K_3 w, t \rightarrow \frac{1}{a_2} t, x \rightarrow \sqrt{\frac{\lambda_2}{a_2}} x$$

увівши нові позначення:

$$\alpha_1 = \frac{c_{31}K_3}{a_2}, \alpha_2 = \frac{c_{12}K_2}{a_2}, \alpha_3 = \frac{c_{13}K_1}{a_2}, \alpha_4 = \frac{c_{32}K_2}{a_2},$$

$$\beta_1 = \frac{a_1}{a_2}, \beta_3 = \frac{a_3}{a_2}, K_1 = \frac{K_3}{K_2}, K_2 = \frac{K_1}{K_2}, d_1 = \frac{\lambda_1}{\lambda_2}, d_3 = \frac{\lambda_3}{\lambda_2}$$

І маємо наступну еквівалентну форму:

$$\begin{aligned} u_t &= d_1 u_{xx} + \beta_1 u (1 - u) - \alpha_1 u w + \alpha_2 u v, \\ v_t &= v_{xx} + v (1 - v) + (K_1 \alpha_3 + K_2 \alpha_1) u w - (K_2 \alpha_2 u + K_1 \alpha_4 w) v, \\ w_t &= d_3 w_{xx} + \beta_3 w (1 - w) - \alpha_3 u w + \alpha_4 v w. \end{aligned} \quad (3)$$

Отримана система з $\alpha_1 = \alpha_3 = 0$ є частковим випадком відомого DLVS, який описує велику кількість процесів у біології та хімії. Однак, вищезазначене обмеження еквівалентно $c_{13} = c_{31} = 0$ у (2), що суперечить основним обмеженням у моделі. Таким чином, вважатимемо, що $c_{13}^2 + c_{31}^2 \neq 0 \Leftrightarrow \alpha_1^2 + \alpha_3^2 \neq 0$, тобто отримана система не є еквівалентною DLVS [9].

1.3 Висновки за розділом 1

Серед великого різноманіття моделей необхідно підібрати ту, яка враховує якомога більше особливостей властивостей даного процесу. Тому, розглянувши систему «хижак-жертва», у якій використовується два диференціальні рівняння, і яка описує кінетику чисельності популяції із одним типом хижаків і одним типом жертв та трикомпонентну нелінійну математичну модель, яка описує взаємодію спільноти, у якій індивіди розмовляють декількома мовами та деколи переходять із рідної мови на користь іншої та навпаки – надалі у роботі буде розглянуто саме реакційно-дифузійну систему, за допомогою якої можна отримати більш точні розв'язки.

РОЗДІЛ 2 ЗАСТОСУВАННЯ МОДЕЛЕЙ ДО ПРОЦЕСУ ЗМІНИ КІЛЬКОСТІ НОСІЇВ МОВИ В УКРАЇНІ

2.1 Застосування системи Лотки-Вольтерри

Вся інформація про чисельність і склад населення країни в цілому включає характеристики його територіального розміщення, віково-статевого і національного складу, сімейно-шлюбної структури, рівня освіти, розподілу за видами занять і джерелом засобів існування та ряд інших даних. Як користувачів інформації про населення виступають практично всі категорії суб'єктів суспільних відносин: органи державної влади та місцевого самоврядування, всі групи юридичних осіб та фізичні особи, а також міжнародні органи.

Офіційна статистична інформація про чисельність і склад населення відповідає загальноприйнятим критеріям якості.

У міру збільшення періоду при розрахунках накопичується похибка розрахункових даних про чисельність населення [5].

В даному випадку, будемо вважати “хижаком” кількість україномовних носіїв. Для застосування цієї моделі нам потрібно визначити константи, що буде показано далі.

Сформулюємо модель в середовищі Excel:

E	F	G	H	I	J	K	L
Константи							
r_1 - народжуваність жертви							2
p_1 - коефіцієнт хижацтва для жертви							0
d_2 - смертність хижака							10
p_2 - коефіцієнт хижацтва							0

Рисунок 2.1.1 Позначення констант

Для першого покоління впишемо певне співвідношення: 210900 жертв та 40993000 хижаків:

	A	B	C
1	Покоління	Хижак\Рівень смертності	Жертва\Чисельність населення в Україні
2	1	210900	40993000

Рисунок 2.1.2 Початкові данні

Надалі в друге і наступні покоління (розрахунок проводився для 200 поколінь) вводимо наступні формули:

	A	B	C
1	Покоління	Хижак\Рівень смертності	Жертва\Чисельність населення в Україні
2	1	210900	40993000
3	2	$=B2+H\$7*B2*C2-G\$7*B2$	$=C2+E\$7*C2-F\$7*C2*B2$
4	3	$=B3+H\$7*B3*C3-G\$7*B3$	$=C3+E\$7*C3-F\$7*C3*B3$
5	4	$=B4+H\$7*B4*C4-G\$7*B4$	$=C4+E\$7*C4-F\$7*C4*B4$
6	5	$=B5+H\$7*B5*C5-G\$7*B5$	$=C5+E\$7*C5-F\$7*C5*B5$

Рис 2.1.3 Формульний вигляд таблиці

	A	B	C
1	Покоління	Хижак\Рівень смертності	Жертва\Чисельність населення в Україні
2	1	210900	40993000
3	2	208791	41812860
4	3	206703	42649117
5	4	204636	43502100
6	5	202590	44372142

Рисунок 2.1.4 Чисельні данні

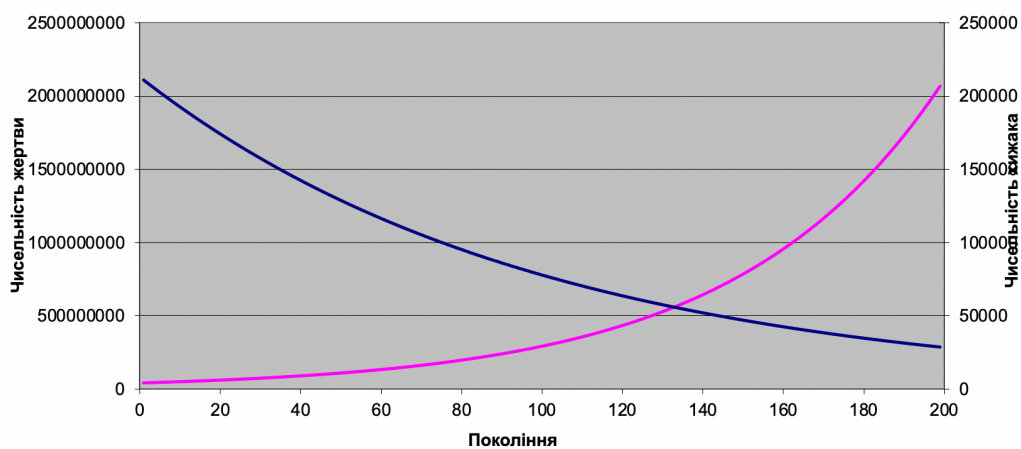


Рисунок 2.1.5 Графічне зображення зростання кількості україномовного населення

Із рисунку 1.5 можна побачити, що загальна чисельність україномовного населення зберігає тенденцію збільшення. Також, природний приріст має позитивне значення, що свідчить про те, що народжуваність населення перевищує його смертність. Але такий приріст україномовного населення може бути зумовлений не лише через загальне демографічне збільшення населення, а й через вплив низки різних соціальних факторів.

Оскільки феномен є контекстуально залежним, при математичному моделюванні подібних динамічних процесів необхідно враховувати вплив багатьох чинників. У зв'язку із цим, дана модель для аналізу та подальшого прогнозування зміни кількості носіїв української(та інших) мов в Україні є не зовсім коректною та грубою.

2.2 Застосування трикомпонентної нелінійної математичної моделі.

Підготовка даних

Скориставшись офіційними даними [2,3,7,8], підготуємо Excel файл для подальшої роботи з ним (жовтим кольором виділені найбільш виражені зміни, які можна проінтерпретувати як історичні події, що мали потенційний вплив на феномен).

Загальна сума відсотків може не дорівнювати 100, оскільки є респонденти, які відмовляються відповідати на питання, або які зазначають у відповідях варіант «інше». Як правило, така кількість відповідей від респондентів не перевищує 1%, що є в межах статистичної похибки. На вибірку національного опитування в Україні, стандартна статистична рівня 3,5%.

Маючи офіційні данні по загальній кількості населення в Україні(population) [8]; відсотку тих, хто говорить лише українською(ukr);

відсотку тих, хто говорить лише російською (ru); відсотку тих, хто говорить в рівній мірі українською та російською (both):

year	population	ukr	ru	both
1994	52114400	36,70%	32,40%	29,40%
1996	51297100	36,90%	33,10%	29,60%
1998	50370800	37,60%	33,40%	28,40%
2000	49429800	39,10%	36,00%	24,80%
2002	48457100	38,20%	33,20%	28,00%
2004	47622400	38,40%	34,30%	26,30%
2005	47280800	41,80%	36,40%	21,60%
2006	46929500	38,00%	39,20%	22,60%
2008	46372700	42,00%	36,80%	19,90%
2010	45962900	41,70%	34,90%	22,10%
2011	45778500	42,80%	38,60%	17,10%
2012	45633600	42,90%	35,40%	20,90%
2013	45553000	38,70%	37,90%	22,80%
2014	45426200	43,20%	31,20%	24,70%
2015	42928900	47,00%	32,00%	20,00%
2016	42760500	47,00%	32,00%	20,00%
2017	42584500	46,00%	32,00%	20,00%
2018	42386400	49,00%	27,00%	23,00%
2019	42153200	51,00%	27,00%	21,00%
2020	41902400	49,00%	25,00%	25,00%

Рисунок 2.2.1 Початкові офіційні данні

Попередньо підготуємо данні та знайдемо прискорення зміни приросту у конкретний момент часу(OLD_uhx, OLD_wxx, OLD_vxx):

СЧЕТЕСЛИ ✖ ✓ f_x =C4+C2-2*C3								
	A	B	C	D	E	F	G	H
1	year	population	ukr	ru	both	OLD_uxx	OLD_wxx	OLD_vxx
2	1994	52114400	36,70%	32,40%	29,40%			
3	1996	51297100	36,90%	33,10%	29,60%			
4	1998	50370800	37,60%	33,40%	28,40%	=C4+C2-2*C3	=D4+D2-2*D3	=E4+E2-2*E3
5	2000	49429800	39,10%	36,00%	24,80%	0,80%	2,30%	-2,40%
6	2002	48457100	38,20%	33,20%	28,00%	-2,40%	-5,40%	6,80%
7	2004	47622400	38,40%	34,30%	26,30%	1,10%	3,90%	-4,90%
8	2005	47280800	41,80%	36,40%	21,60%	3,20%	1,00%	-3,00%
9	2006	46929500	38,00%	39,20%	22,60%	-7,20%	0,70%	5,70%
10	2008	46372700	42,00%	36,80%	19,90%	7,80%	-5,20%	-3,70%
11	2010	45962900	41,70%	34,90%	22,10%	-4,30%	0,50%	4,90%
12	2011	45778500	42,80%	38,60%	17,10%	1,40%	5,60%	-7,20%
13	2012	45633600	42,90%	35,40%	20,90%	-1,00%	-6,90%	8,80%
14	2013	45553000	38,70%	37,90%	22,80%	-4,30%	5,70%	-1,90%
15	2014	45426200	43,20%	31,20%	24,70%	8,70%	-9,20%	0,00%
16	2015	42928900	47,00%	32,00%	20,00%	-0,70%	7,50%	-6,60%
17	2016	42760500	47,00%	32,00%	20,00%	-3,80%	-0,80%	4,70%
18	2017	42584500	46,00%	32,00%	20,00%	-1,00%	0,00%	0,00%
19	2018	42386400	49,00%	27,00%	23,00%	4,00%	-5,00%	3,00%
20	2019	42153200	51,00%	27,00%	21,00%	-1,00%	5,00%	-5,00%
21	2020	41902400	49,00%	25,00%	25,00%	-4,00%	-2,00%	6,00%

Рисунок 2.2.2 Попередня підготовка даних

Далі знаходимо середню зміну приросту, оскільки такі данні є непередбачуваними, а середнє значення показує середню тенденцію щодо зростання або спадання(uxx, wxx, vxx):

CP3HAЧ ✖ ✓ f_x =CP3HAЧ(F\$4:F4)											
	A	B	C	D	E	F	G	H	I	J	K
1	year	population	ukr	ru	both	OLD_uxx	OLD_wxx	OLD_vxx	uxx	wxx	vxx
2	1994	52114400	36,70%	32,40%	29,40%						
3	1996	51297100	36,90%	33,10%	29,60%						
4	1998	50370800	37,60%	33,40%	28,40%	0,50%	-0,40%	-1,40%	=CP3HAЧ(F\$4:F4)	=CP3HAЧ(G\$4:G4)	=CP3HAЧ(H\$4:H4)
5	2000	49429800	39,10%	36,00%	24,80%	0,80%	2,30%	-2,40%	0,65%	0,95%	-1,90%
6	2002	48457100	38,20%	33,20%	28,00%	-2,40%	-5,40%	6,80%	-0,37%	-1,17%	1,00%
7	2004	47622400	38,40%	34,30%	26,30%	1,10%	3,90%	-4,90%	0,00%	0,10%	-0,48%
8	2005	47280800	41,80%	36,40%	21,60%	3,20%	1,00%	-3,00%	0,64%	0,28%	-0,98%
9	2006	46929500	38,00%	39,20%	22,60%	-7,20%	0,70%	5,70%	-0,67%	0,35%	0,13%
10	2008	46372700	42,00%	36,80%	19,90%	7,80%	-5,20%	-3,70%	0,54%	-0,44%	-0,41%
11	2010	45962900	41,70%	34,90%	22,10%	-4,30%	0,50%	4,90%	-0,06%	-0,33%	0,25%
12	2011	45778500	42,80%	38,60%	17,10%	1,40%	5,60%	-7,20%	0,10%	0,33%	-0,58%
13	2012	45633600	42,90%	35,40%	20,90%	-1,00%	-6,90%	8,80%	-0,01%	-0,39%	0,36%
14	2013	45553000	38,70%	37,90%	22,80%	-4,30%	5,70%	-1,90%	-0,40%	0,16%	0,15%
15	2014	45426200	43,20%	31,20%	24,70%	8,70%	-9,20%	0,00%	0,36%	-0,62%	0,14%
16	2015	42928900	47,00%	32,00%	20,00%	-0,70%	7,50%	-6,60%	0,28%	0,01%	-0,38%
17	2016	42760500	47,00%	32,00%	20,00%	-3,80%	-0,80%	4,70%	-0,01%	-0,05%	-0,01%
18	2017	42584500	46,00%	32,00%	20,00%	-1,00%	0,00%	0,00%	-0,08%	-0,05%	-0,01%
19	2018	42386400	49,00%	27,00%	23,00%	4,00%	-5,00%	3,00%	0,18%	-0,36%	0,18%
20	2019	42153200	51,00%	27,00%	21,00%	-1,00%	5,00%	-5,00%	0,11%	-0,04%	-0,13%
21	2020	41902400	49,00%	25,00%	25,00%	-4,00%	-2,00%	6,00%	-0,12%	-0,15%	0,21%

Рисунок 2.2.3 Підготовка даних

2.3 Програмна реалізація аналогу реакційно-дифузійної системи за допомогою Python

Розглянемо кожну із операцій за допомогою Python.

- 1) Імпортуємо необхідні бібліотеки для роботи із даними та готуємо трикомпонентну нелінійну модель згідно підрозділу 1.2

▼ Imports

```
[ ] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

▼ Prepare math model

```
|
def u(x,t):
    return float(data[data["year"] == t]['ukr'])

def v(x,t):
    return float(data[data["year"] == t]['both'])

def w(x,t):
    return float(data[data["year"] == t]['ru'])

def uxx(x, t):
    return float(data[data["year"] == t]['uxx'] )

def vxx(x, t):
    return float(data[data["year"] == t]['vxx'])

def wxx(x, t):
    return float(data[data["year"] == t]['wxx'])

def K():
    return data['population'].max()

def x(t):
    return data[data.year == t]['population']
```

```

def udt(x, t):
    res = uxx(x,t) + u(x,t)*(1-(u(x,t)/(K() - (v(x,t)+w(x,t)))))*0 - c31*u(x,t)*w(x,t)+c12*u(x,t)*v(x,t)
    return float(res)

def vdt(x, t):
    res = vxx(x,t) + v(x,t)*(1-(v(x,t)/(K() - (u(x,t)+w(x,t)))))*0 + (c13 + c31) * u(x,t)*w(x,t) - (c12*u(x,t) + c32*w(x,t))*v(x,t)
    return float(res)

def wdt(x, t):
    res = wxx(x,t) + w(x,t)*(1-(w(x,t)/(K() - (v(x,t)+u(x,t)))))*0 - c13*u(x,t)*w(x,t)+c32*v(x,t)*w(x,t)
    return float(res)

```

Рисунок 2.3.1 Результат підготовки математичної моделі

2) Введемо значення коефіцієнтів (прим. c_{31} та $c_{13} = 0,5$, оскільки у соціальних опитуваннях варіант відповіді на питання про мову звучить «рівною мірою спілкуюсь українською та російською мовою», значення c_{12} та c_{32} відповідають офіційним даним) та генеруємо прогноз зміни кількості носіїв української мови та інших мов в Україні:

▼ Input coefs and data, generate forecast and plot

```

data = pd.read_excel("data.xlsx")

c13 = 0.5 # ->bi->ukr
c31 = 0.5 # ->bi->ru

c12 = 0.73 # ->ukr
c32 = 0.26 # ->ru OFFICIAL
#probabilities to change language
years = list(range(2021,2030))

uxx_future = data.uxx.mean()
vxx_future = data.vxx.mean()
wxx_future = data.wxx.mean()

for y in years:
    data = data.append(pd.Series([int(y),0,u(0,y-1)+udt(0,y-1),v(0,y-1)+vdt(0,y-1),
    w(0,y-1)+wdt(0,y-1),uxx_future,vxx_future,wxx_future],index=data.columns), ignore_index=True)

```

Рисунок 2.3.2. Прогнозування та введення даних

3) Та зображення графіку зміни кількості носіїв мови:

```

data.plot(x="year",y=["ukr", 'ru', "both"],kind="line")
plt.show()

```

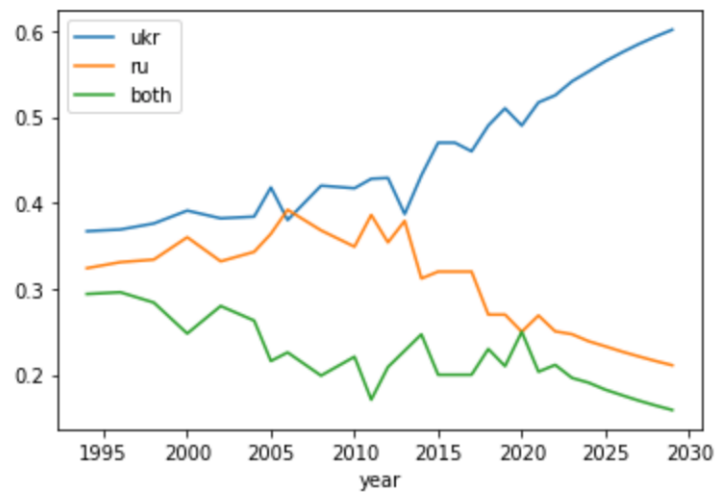


Рисунок 2.3.3. Результат у графічному вигляді при значенні коефіцієнтів c_{31} та $c_{13} = 0,5$ (ймовірність переходу на іншу мову з рідної)

Також, розглянемо випадок, коли ймовірність переходу на іншу мову з рідної $c_{31} = 0,64$ та $c_{13} = 0,36$.

```

data = pd.read_excel("data.xlsx")

c13 = 0.36 # ->bi->ukr
c31 = 0.64 # ->bi->ru

c12 = 0.73 # ->ukr
c32 = 0.26 # ->ru OFFICIAL
#probabilities to change language
years = list(range(2021,2030))

uxx_future = data.uxx.mean()
vxx_future = data.vxx.mean()
wxx_future = data.wxx.mean()
for y in years:
    data = data.append(pd.Series([int(y),0,u(0,y-1)+udt(0,y-1),v(0,y-1)+vdt(0,y-1),
    w(0,y-1)+wdt(0,y-1),uxx_future,vxx_future,wxx_future],index=data.columns), ignore_index=True)

data.plot(x="year",y=["ukr", 'ru', "both"],kind="line")
plt.show()

```

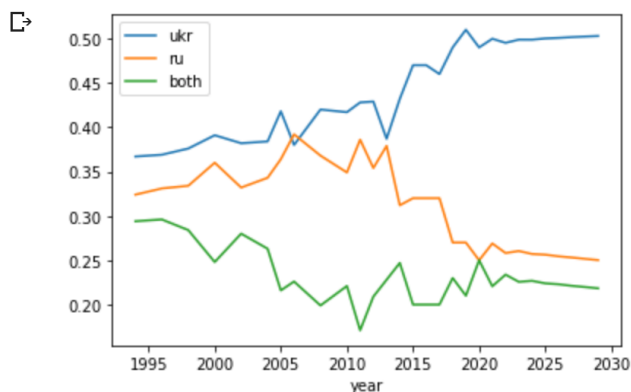


Рисунок 2.3.4 Результат у графічному вигляді при значенні коефіцієнтів (ймовірність переходу на іншу мову з рідної) при $c_{31}=0,64$ та $c_{13} = 0,36$

З рисунку 2.3.4 можна побачити, що для зростання кількості україномовних носіїв необхідна ймовірність щонайменше 0,36% задля переваги української мови над російською. В протилежному випадку можна спостерігати спад кількості україномовних носіїв:

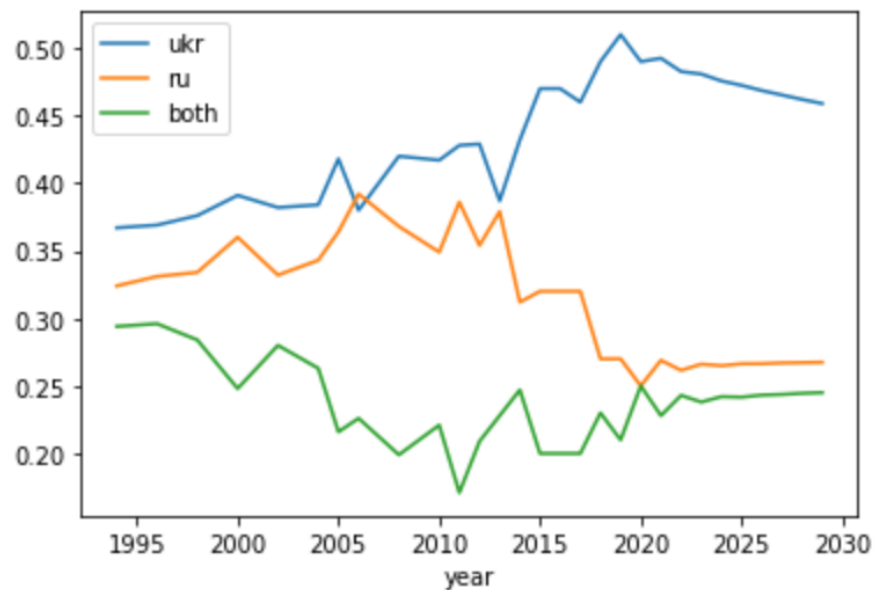


Рисунок 2.3.5 Результат у графічному вигляді при значенні коефіцієнтів (ймовірність переходу на іншу мову з рідної) при $c_{31}=0,7$ та $c_{13} = 0,3$

Результати, отримані за моделлю, описану в підрозділі 1.2 демонструють загальну тенденцію помірного зростання кількості україномовних носіїв за ідеальних умов. Модель прогнозує загальну картину, не беручи до уваги можливість впливу соціальних факторів, які можуть виникати впродовж даного періоду.

Підсумовуючи, можна сказати, що із 1995 року спостерігається плавне але стабільне збільшення кількості україномовного населення, в той час, коли людей, які ідентифікують себе як двомовні стабільно стає менше. В період із 1995 по 2014 рік, відсоток людей, що ідентифікують себе як російськомовні коливався в межах 30-40%, в той час як після 2014 року ми можемо спостерігати стрімке зниження їхнього відсотку. Можна навести багато потенційних інтерпретацій того, чому так відбулося, зокрема, втрата Україною територій із найбільшим відсотком російськомовного населення (Крим та окремі регіони Донецької та Луганської областей), зміна суспільно-політичної орієнтації в країні, чи навіть впровадження закону про мову, що вплинув на

суспільні настрої. Інтерпретація наведених соціальних явищ та визначення причин зменшення кількості російськомовного населення не є ціллю даного дослідження. Проте можна з точністю запевняти, що дана реакційно-дифузійна система демонструє стабільне збільшення кількості носіїв української мови.

ВИСНОВКИ

В роботі досліджено і проаналізовано дві математичні моделі, та сконцентровано увагу на модель, яка спрямована на визначення тенденції зміни кількості носіїв мови. Підготовлено офіційні данні для подальшого використання та за допомогою Python спрогнозовано кількість носіїв мови в Україні до 2030 року.

Результати, отримані за моделлю Лотки-Вольтерри враховують недостатньо факторів для подальшого прогнозування тенденції зміни носіїв мови в Україні та враховуючи, що феномен є контекстуально залежним, при математичному моделюванні подібних динамічних процесів необхідно враховувати вплив багатьох чинників. У зв'язку із цим, дана модель саме для аналізу та подальшого прогнозування зміни кількості носіїв української(та інших) мов в Україні є не зовсім коректною.

Розглянувши трикомпонентну нелінійну модель було визначено, що вона дає змогу спрогнозувати і побачити загальну картину не беручи до уваги можливість впливу соціальних факторів, які можуть виникати впродовж даного періоду. Враховуючи результати, отримані за допомогою аналогу реакційно-дифузійної системи можна підсумувати, що загальна тенденція збільшення носіїв української мови переважає, оскільки для того, щоб дана кількість носіїв зросла потрібен менший відсоток ймовірності переходу із російської мови на українську($c_{13} = 0,36\%$).

Елементом новизни є розгляд та застосування моделі [9,10] на офіційних даних, та розробка програмного коду на Python.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Валуєвський циркуляр [Електронний ресурс] – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/Валуєвський_циркуляр.
2. Думки та погляди населення України: Серпень 2020 [Електронний ресурс] // Центр соціальний моніторинг. – 2020. – Режим доступу до ресурсу: <https://smc.org.ua/dumky-ta-poglyady-naselennya-ukrayiny-serpen-2020-1463/>.
3. ІНФОРМАЦІЙНО-АНАЛІТИЧНІ МАТЕРІАЛИ З ПИТАНЬ ДЕРЖАВНОЇ МОВНОЇ ПОЛІТИКИ [Електронний ресурс] // Міністерство культури України. – 2015. – Режим доступу до ресурсу: http://mincult.kmu.gov.ua/control/publish/article?art_id=244971395.
4. Модель Лотки-Вольтерры [Електронний ресурс] – Режим доступу до ресурсу: https://math-it.petrus.ru/users/semenova/MathECO/Lectons/Lotka_Volterra.pdf.
5. Про затвердження методологічних рекомендацій з питань статистики населення [Електронний ресурс] // Державний комітет статистики України – Режим доступу до ресурсу: <https://ips.ligazakon.net/document/FIN36826>.
6. Рівняння Лотки-Вольтерри [Електронний ресурс] – Режим доступу до ресурсу: 1. https://uk.wikipedia.org/wiki/Рівняння_Лотки_—_Вольтерри.
7. Українська мова: шлях у незалежній Україні [Електронний ресурс] // Фонд демократичні ініціативи. – 2020. – Режим доступу до ресурсу: <https://dif.org.ua/article/ukrainska-mova-shlyakh-u-nezalezhniy-ukraini>.
8. Чисельність населення та середня чисельність за періоди року [Електронний ресурс] // Державна служба статистики України – Режим доступу до ресурсу: <http://www.ukrstat.gov.ua>.

9. Cherniha R. Exact Solutions of a Mathematical Model Describing Competition and Co-Existence of Different Language Speakers / Cherniha Roman, 2020. – 11 с.
10. Kandler A. Language shift, bilingualism and the future of Britain's Celtic languages / A. Kandler, R. Unger, J. Steele. // the royal society publishing. – 2010.
11. Lotka A. J. Elements of physical biology. Baltimore: Williams and Wilkins Co., 1925.
12. What is mathematical modelling [Электронный ресурс] // Australian Council for Educational Research International Mathematical Modeling Challenge. – 2007. <https://www.immchallenge.org.au/supporting-resources/what-is-mathematical-modelling>.

ДОДАТКИ

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

def u(x,t):
    return float(data[data["year"] == t]['ukr'])

def v(x,t):
    return float(data[data["year"] == t]['both'])

def w(x,t):
    return float(data[data["year"] == t]['ru'])

def uxx(x, t):
    return float(data[data["year"] == t]['uxx'] )

def vxx(x, t):
    return float(data[data["year"] == t]['vxx'])

def wxx(x, t):
    return float(data[data["year"] == t]['wxx'])

def K():
    return data['population'].max()

def x(t):
    return data[data.year == t]['population']

def udt(x, t):
    res = uxx(x,t) + u(x,t)*(1-(u(x,t)/(K() - (v(x,t)+w(x,t)))))*0 -
        c31*u(x,t)*w(x,t)+c12*u(x,t)*v(x,t)
    return float(res)

def vdt(x, t):
    res = vxx(x,t) + v(x,t)*(1-(v(x,t)/(K() - (u(x,t)+w(x,t)))))*0 + (c13 +
        c31) * u(x,t)*w(x,t) - (c12*u(x,t) + c32*w(x,t))*v(x,t)
    return float(res)

def wdt(x, t):
    res = wxx(x,t) + w(x,t)*(1-(w(x,t)/(K() - (v(x,t)+u(x,t)))))*0 -
        c13*u(x,t)*w(x,t)+c32*v(x,t)*w(x,t)
    return float(res)
data = pd.read_excel("data.xlsx")

```

```

c13 = 0.36 # ->bi->ukr
c31 = 0.64 # ->bi->ru
c12 = 0.73 # ->ukr
c32 = 0.26 # ->ru OFFICIAL
#probabilities to change language
years = list(range(2021,2030))

uxx_future = data.uxx.mean()
vxx_future = data.vxx.mean()
wxx_future = data.wxx.mean()

for y in years:
    data = data.append(pd.Series([int(y),0,u(0,y-1)+udt(0,y-1),v(0,y-1)+vdt(0,y-1),
    w(0,y-1)+wdt(0,y-1),uxx_future,vxx_future,wxx_future],index=data.columns),
        ignore_index=True)

data.plot(x="year",y=["ukr", 'ru', "both"],kind="line")
plt.show()

```