

COOPERATIVE AND CONFLICT COMMUNICATION: GENERALIZING OF THE THEORETIC BASIS

The article summarizes various opinions on cooperative and conflict communication and linguistical and extra-linguistical factors which affect successful communicative process. In the context of typology of people interaction and in the light of communicative linguistics we generalize such concepts as cooperation and conflict, communicative goal, strategy and tactics, communicative personality, social role and status of speakers.

Keywords: cooperative and conflict communication, communication, conflict, cooperation, typology of people interaction.

УДК 811.161.2'42:004

Мяснікова В. В., магістрант,
Інститут філології КНУ імені Тараса Шевченка, м. Київ

АВТОМАТИЧНА ФОНЕТИЧНА ТРАНСКРИПЦІЯ ТА ТРАНСЛІТЕРАЦІЯ УКРАЇНСЬКОГО ТЕСТУ

У статті комплексно розглянуто програму автоматичної фонетичної транскрипції та транслітерації українського тексту «TranSScripT». Особливу увагу приділено принципам роботи програми, яка забезпечує автоматичну фонетичну транскрипцію, транслітерацію та зворотну транслітерацію текстових масивів. Наголошено на унікальності дослідження та його теоретичному і прикладному значенні.

Ключові слова: автоматичне опрацювання тексту, алгоритм, транслітерація, фонетична транскрипція.

З прогресом у галузі інформаційно-комунікаційних технологій розвиваються і технології опрацювання мовлення, які реалізують пошуки у розв'язанні класичних проблем діалогу «людина – машина». У світовому науковому просторі такі технології часто спрямовані на широке коло користувачів і тому є популярними.

Такими є продукти технологій опрацювання усного (розпізнавання природного мовлення – Speech Recognition, синтез штучного мовлення – Speech Synthesis) та писемного мовлення (корпусна лінгвістика).

Аспект усномовленнєвого опрацювання втілюється, відповідно до розвитку персональних мобільних систем, у програмах голосового управління мобільними пристроями та віртуальних асистентах: *Siri* – помічник, розроблений для операційної системи iOS; *Microsoft Cortana* – помічник для пристроїв на платформі Windows Phone; *Google Now* – помічник від Google, *S Voice* – від Samsung). Усі вони працюють

на основі синтезатора штучного мовлення і є спрямованими на розпізнавання людського мовлення та адекватне реагування на нього у режимі реального часу: відповіді на запитання, пошук інформації, зміна налаштувань, запуск програмного забезпечення тощо. Зворотним процесом до розпізнавання мовлення є синтез мовлення, який досить успішно реалізується на базі синтезаторів мовлення Text-to-Speech, активно розроблюваних світовими ІТ-компаніями.

Другий аспект – опрацювання писемного мовлення (власне тексту) – забезпечується масивами текстів, укладеними у корпус. «Корпусні студії зосереджуються на аналізі природної мови в умовах реального функціонування з використанням комп'ютерних технологій на основі великих за обсягом, ретельно відібраних та впорядкованих текстових корпусів» [6, 548].

Корпус – «електронне зібрання текстів природної мови, впорядковане, організоване й оформлене певним чином, призначене для наукового та практичного вивчення мови» [1, 2]. Окрім власне текстових масивів він містить метатекстову інформацію, т. зв. розмітку, яка передбачає, що до звичного у нашому розумінні тексту додається певна формальна інформація шляхом вписування у класичний текст відповідних символів, за допомогою яких машина розуміє цей текст і може відповісти на питання щодо нього [3, 35]. Такою розміткою є морфологічне та синтаксичне анотування корпусу.

Мета нашого дослідження – створення унікального програмного забезпечення для автоматичної фонетичної транскрипції та транслітерації українськомовного тексту – є перспективною з урахуванням обох аспектів сучасних прикладних лінгвістичних досліджень.

Актуальність теми безпосередньо пов'язана з тим фактом, що на сучасному етапі автоматичного опрацювання тексту необхідним є створення системи автоматичної фонетичної транскрипції для забезпечення фонетичного опрацювання тексту на рівні з автоматичним морфологічним, синтаксичним та семантичним аналізом тексту. Не менш важливим постає використання алгоритму автоматичної транскрипції у модулях лінгвістичного аналізу синтезаторів штучного мовлення та пов'язаних з ними програмних додатків, призначених для розв'язання лінгвістичних завдань.

Програма «TranScripT» реалізована в інтегрованому середовищі розроблення з відкритим вихідним кодом NetBeans IDE (Integrated Development Environment) на платформі Java. «TranScripT» трансліюється в байт-код, що виконується віртуальною машиною Java (Java Virtual Machine) – програмою, що обробляє байтовий код, передає інструкції обладнанню і виступає інтерпретатором. Перевагою такого способу виконання програми є повна незалежність байт-коду від операційної системи та обладнання, що дозволяє створювати та використовувати Java-додатки на будь-якому пристрої, для якого існує відповідна віртуальна машина [6, 200]. Саме тому наша програма має можливість застосування не лише на стандартних та портативних персональних комп'ютерах за допомогою платформи Java, а і на телефонах, смартфонах, планшетах та інших пристроях, які підтримують цю технологію. У перспективі можливим є перенесення функціоналу програми на спеціальний веб-ресурс, що дозволить здійснювати фонетичну транскрипцію та транслітерацію за допомогою мережі Інтернет без встановлення програмного забезпечення на комп'ютер.

Графічний інтерфейс програми автоматичної транскрипції та транслітерації створено за допомогою бібліотеки JavaSwing, тому «TranSScripT» має вигляд додатка, типового для використовуваної операційної системи.

Головне вікно програми містить стандартні компоненти, такі як панель меню, кнопки згортання, регулювання розміру вікна та закриття програми.

Меню «Файл» дозволяє закрити програму; «Правлення» – копіювати, вирізати, вставити, виділити весь текст або текст окремих полів; «Довідка» – викликати довідку, яка містить коротку інструкцію для користувача та дані для зв'язку з розробниками програми.

Основними компонентами програми є поля введення тексту та виводу результату, кнопки «Результат», «Транскрипція», «Транслітерація», «Зворотна транслітерація», «Очистити». Остання використовується для очищення усіх полів (Рис. 1. Інтерфейс програми автоматичної фонетичної транскрипції та транслітерації українського тексту «TranSScripT»).

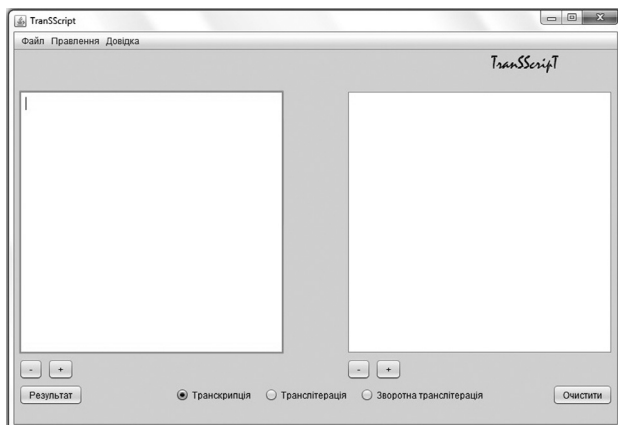


Рис. 1. Інтерфейс програми автоматичної фонетичної транскрипції та транслітерації українського тексту «TranSScripT».

Представлена нами програма містить широкі поля введення і виведення даних та смуги прокручування, за допомогою яких здійснюється навігація по полям при великому об'ємі тексту. Також у вікні наявні кнопки зменшення і збільшення розміру шрифту, який використовується у відображенні тексту, що робить користування програмою винятково зручним: великий текст можна зменшити для відображення максимальної кількості рядків без прокручування; маленький текст можна збільшити для комфортного перегляду.

«TranSScripT» здатний опрацювати 2147483647 символів, що, однак, не може не впливати на швидкість опрацювання тексту. Тому наразі нами проводиться процес перебудови програми та оптимізації її коду.

У поле введення може бути внесений лише український текст, що забезпечує автоматичну фонетичну транскрипцію та транслітерацію, та транслітерований латиницею український текст для зворотної транслітерації.

При вставленні або наборі тексту у поле введення користувач обирає модуль, у межах якого відбуватиметься опрацювання тексту («Транскрипція», «Транслітерація», «Зворотна транслітерація»). При натисненні кнопки «Результат» програма розбиває слова тексту на масиви. Формальне сприйняття тексту комп'ютером передбачає розуміння слова як сукупності графічних символів між пробілами.

Головний цикл програми аналізує оточення елементів масиву, порівнює кожний символ з попереднім, наступним і будь-яким символом, заданим алгоритмом на основі правил, та відповідно до них формує масив даних, який виводиться у поле результату.

За бажанням користувача програма може або очистити поля для введення іншого масиву тексту та, відповідно, нового циклу програми, або скопіювати результат за допомогою меню «Правлення»/контекстного меню у файл Microsoft Word. Застереженням при роботі з текстовим редактором є підтримка фонетичного шрифту ZRCola для повного відображення використовуваних символів.

На вибір користувача програма може здійснити транскрипцію, транслітерацію (українська-латинка) та зворотну транслітерацію (латинка-українська) українського тексту.

Фонетичний модуль програми «TranSScripT» є системою, на вхід якої подається графічний текст, що містить виключно символи алфавіту української мови та графічний символ межі між словами – пробіл. Опрацювання текстів проводиться автоматично, а це передбачає переведення тексту у фонетичне представлення за відповідними правилами транскрипції. Програма «TranSScripT» працює виключно з фонетичною транскрипцією, метою якої є «<...> якомога точніше передати вимову звуків з усіма її відтінками» [5, 174].

Наразі програма працює з такими змінами звуків, як:

1) акомодация:

а) лабіалізація;

б) назалізація;

в) палаталізація;

г) і-подібна артикуляція;

2) асиміляція:

а) за місцем творення;

в) за місцем і способом творення.

Вони алгоритмізуються за такими правилами:

1. Програма, здійснюючи дистрибутивний аналіз одиниць слова, визначає відповідності між літерами, комбінаціями літер та їх фонетичними еквівалентами.

2. Кожному звукові, позначеному на письмі, відповідає літера транскрипції з діакритичними знаками.

3. Запис береться у квадратні дужки: [поле].

4. Великі літери автоматично переводяться у прописні: *Марина* – [марина].

5. Пунктуаційні знаки речень передаються через позначки паузації»/» (для»,»), «//» (для»,»), «!», «?», «;», «:») та»//» (для «...», «?!», «?!», «!!!», «?!..» «!..»).

6. М'якість попереднього приголосного позначається як «[˘]», пом'якшеність – «[˙]», подовження – «:».

7. М'який знак, літери «я»; «ю»; «є»; «і» позначають м'якість попереднього приголосного і у транскрипції визначають його діакритичні знаки.

7.1. Букви «я», «ю», «є» можуть позначати як один, так і два звуки. Два звуки «я», «ю», «є» позначають лише у трьох випадках:

1) коли вони стоять на початку слова: *яблуко* – [йяблуко]; *Єреван* – [йереван];

2) в середині слова після голосного: *маяк* – [майак], *Аюдаг* – [айудаг];

3) після апострофа і м'якого знака я, ю, є: *здоров'я* – [здоров'я], *Мольєр* – [мол'йєр].

7.2. У випадку з пом'якшенням приголосних, яке позначають на письмі букви «я», «ю», «є», «і», у транскрипції на м'якість або пом'якшеність вказують діакритичний знак м'якості [˘] або пом'якшеності [˙]. Стосуються вони приголосного звука: *малюють* – [мал'уйут˘].

7.3. До м'яких приголосних належать [д˘], [т˘], [з˘], [с˘], [ц˘], [л˘], [н˘], [д̂з˘], [р˘], якщо вони стоять перед «я», «ю», «є», «і», та [й]. Останній є виключно м'яким і не потребує діакритичного знаку м'якості.

7.4. Напівпом'якшеними стають: губні: [б˘], [п˘], [в˘], [м˘], [ф˘], глотковий [г˘], задньоязикові [х˘], [г˘], [к˘], шиплячі [ж˘], [ш˘], [дж˘], [ч˘] – у випадку, якщо вони стоять перед «я», «ю», «є», «і».

8. «І» та «ц» завжди позначають два звуки: [йі] та [шч].

9. Африкати [д̂з] та [дж] у транскрипції пишуться злитю, над обома частинами африкати розташована з'єднувальна дужка. Однак у випадку зі збігом звуків [д] і [з] або [д] і [ж] на межі префікса і кореня дужка відсутня: [п'ідживити].

10. Буква «в» позначає звук [ʋ̂], а «й» – звук [і] на початку слова перед приголосним: *вчора* – [ʋ̂чора]; у середині слова перед приголосними та «й»; у кінці слова після голосного: *чайний* – [чайній].

11. Носові приголосні [м], [н], [н'] назалізують сусідні голосні як у прогресивному, регресивному, так і в обох напрямках одночасно: *а* – [ã], *о* – [x], *у* – [ʋ̂], *і* – [ĩ], *у* – [ũ], *е* – [ẽ]: *соната* – [сонãта].

12. М'які та пом'якшені приголосні впливають на артикуляцію сусідніх голосних «а», «о», «у», «є», внаслідок чого останні набувають і-подібної артикуляції, що може реалізуватися в прогресивному, регресивному, обох напрямках одночасно: *Катя* – [ка•т'я], *лялька* – [л'а•л'ка].

13. Приголосні звуки української мови регресивно лабіалізуються перед огубленими голосними [о], [у]: *капуста* – [кап°уста], *хуга* – [х°уга], *вухо* – [в°ух°о] [5, с.233-234].

Наприклад, словосполучення «реалізується назалізація» «TranSScripT» затранскрибує таким чином:

Реалізується назалізація →

→ [p][e][a][l][i][z][y][e][tʃ][b][c][jə][_][h][a][z][a][l][i][z][a][n][i][jə] →

→ [p][e][a•][l'ı][z°][y•][ĵe][tʃ˘]:[•a][_][h][ã][z][a•][l'ı][z][a•][ı][ĵə]

→[реа•л'із°у•й•е•ц':•а нāза•л'іза•ц'ій•а] (Рис.2. Робота модуля фонетичної транскрипції).

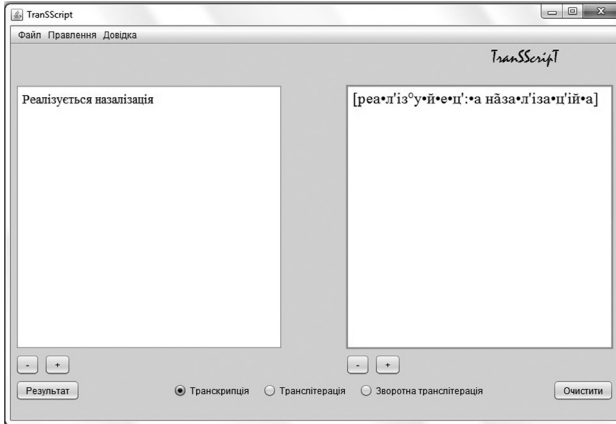


Рис.2. Робота модуля фонетичної транскрипції

Подальшій алгоритмізації підлягають інші види асиміляції приголосних у потоці мовленні (за дзвінкістю/глухістю, м'якістю/твердістю, способом творення) та зміни голосних, які залежать від наголосу.

Найістотношою перешкодою автоматизації транскрипції є рухомий та вільний наголос української мови. Проблема може бути вирішена за допомогою підключення до програми бази даних, яка містить слова з наголосами та наразі розробляється. У даному випадку після введення тексту додатковий цикл програми шукатиме кожне слово у підключеній базі даних, після чого відбуватиметься стандартна процедура транскрипції. У випадку відсутності слова у базі програма запропонує користувачеві поставити наголос самостійно.

Транслітераційний модуль працює за таким же принципом, проте масиви містять не фонетичні символи та діакритичні знаки, а латинські відповідники кириличних літер українського алфавіту, пунктуаційні знаки та знак пробілу.

«TranSScripT» транслітерує словосполучення «великий корпус текстів» так:

Великий корпус текстів→

→[V][e][l][i][k][i][i][][k][o][p][u][s][][t][e][k][c][t][i][v]→

→[V][e][l][i][y][k][y][][k][o][r][p][u][s][][t][e][k][s][t][i][v]→

→*Velykyj korpus tekstiv*

Алгоритм автоматичної транслітерації та зворотної транслітерації, представлений у програмі, є своєрідним підсумком нашої роботи з різними стандартами транслітерації українського алфавіту, якими послуговуються у світі (BGN/PCGN (1965); ТКПН; Українська латинка; Науковий (традиційний); Науковий (ISO/R 9:1968); ISO 9:1995 (ГОСТ 7.79 А); ГОСТ 7.79-2000 Б; УКППТ 1996 (спрощений); ALA-LC;

Паспортний/КМУ 2010). Жодна з цих систем не задовольнила нас в аспектах повної зворотності, простоти символів та сумісності зі стандартним кодуванням.

Нами було внесено повне розрізнення позначень літер *i*, *ï*, *u*, *ï* шляхом введення позначення літери *ï* як *j*. Це допомогло вирішити проблему застосування лише двох латинських для відображення чотирьох українських літер. Також інновація торкнулася відображення м'якого знаку як символу пом'якшення попереднього приголосного: *тінь* – *tin'*, та апострофу, після якого голосні стали йотованими: *polumjia, solovji*.

Зворотна транслітерація створена з метою переведення українського тексту, записаного латиницею, у традиційне кириличне відображення. Її побудовано на основі обернення правил звичайної транслітерації:

Velykyj korpus tekstiv→

→[V][e][l][y][k][y][k][y][j][_][k][o][r][p][u][s][_][t][e][k][s][t][i][v]→

→[B][e][l][i][k][i][i][k][i][i][i][j][_][k][o][p][p][y][c][_][t][e][k][c][t][i][v]→

→*Великий корпус текстів*

Особливість створеної нами системи транслітерації полягає у повній її зворотності та використанні єдиного діакритичного символу (апострофу), що дозволяє уникнути несумісності кодування та неправильного відображення транслітерованого тексту у мережі Інтернет та у текстових редакторах. Згодом наша система транслітерації стане доступною для загального користування на спеціальному сайті www.translitera.ua.

Як будь-яке програмне забезпечення автоматичного опрацювання тексту, програма «TranSScripT» допускає помилки, які мають мовну і технічну природу.

Мовний аспект помилок пов'язаний із тим, що, по-перше, на сучасному етапі комп'ютер не в змозі забезпечити опрацювання людського мовлення так, як може його здійснити досвідчений фахівець-лінгвіст, що є проблемою і галузі штучного інтелекту в цілому; по-друге, мовлення є динамічним процесом, у якому функціонування одних використовуваних одиниць впливає на функціонування інших, що важко формалізувати адекватно.

Технічний аспект помилок виникає як наслідок взаємодії великої кількості умовних операторів програми, на яких вона побудована. Якщо модуль транслітерації та зворотної транслітерації реалізовані найпростішими правилами «якщо..., то...», які передбачають графічний знак на вході і відповідний графічний знак на виході, що виключає різночитання, то модуль транскрипції відображає важко передбачувані для машини комбінації знаків, взаємодія яких у синтагматичному плані може призводити до втрати фонетичних символів.

Так, поширеною помилкою фонетичного модуля «TranSScripT» була втрата символу і-подібної артикуляції, якщо у суміжних символах представлено більше, ніж два діакритичних знаки: [транскриб°ув□н':□]. Як видно з прикладу, наявність чотирьох фонетичних символів підряд спричинило відсутність знаків і-подібної артикуляції перед та після *n* [□н':□]. Наразі проблему вирішено ієрархічною оптимізацією коду.

На наступному етапі ми плануємо запровадити можливість обмеженого ручного редагування тексту, що дозволить вдосконалити його опрацювання.

Теоретичне значення нашого дослідження безпосередньо стосується внеску у дослідження галузі штучного інтелекту, а саме – автоматичного опрацювання тексту, аналізу та синтезу мовлення. **Практичне значення** полягає у тому, що програма «TranSScripT» є надзвичайно корисною для науковців, які займаються проблемами автоматичного опрацювання мовлення та корпусної лінгвістики; лінгвістів при роботі з масивами діалектологічних записів та укладанні «Корпусу діалектних текстів української мови»; викладачів та студентів при вивченні курсу фонетики та суміжних дисциплін. Можемо з упевненістю стверджувати, що наше дослідження виводить можливість комп'ютерного опрацювання українського тексту на новий рівень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Дарчук Н. П. Навчальна програма з дисципліни «Корпусна лінгвістика: проблеми, методи, перспективи»/ Дарчук Н. П. – 2013. – 10 с.
2. Демська-Кульчицька О. М. Базові поняття корпусної лінгвістики / Демська-Кульчицька О. // Українська мова. – 2003. – № 1. – с. 40-45.
3. Демська-Кульчицька О. М. Що таке корпус текстів? / Демська-Кульчицька О. // Культура слова. – 2004. – № 64. – с. 35-38.
4. Демська-Кульчицька О. Основи національного корпусу української мови: монографія / О. Демська-Кульчицька. – К.: ІУМ НАНУ, 2005. – 218 с.
5. Мяснікова В. В., Стенка П. В. Лінгвістичні засади створення алгоритму автоматичної транскрипції і транслітерації українського тексту // *М о в а*, свідомість, художня творчість, інтернет у дзеркалі сучасних філологічних студій. – Київський університет, 2013. – с. 230-237.
6. Шапошникова І. Сучасна українська літературна мова. Фонетика і фонологія. Орфоенія. Графіка і орфографія / Шапошникова І., Гайдученко Г. – Херсон, 2001. – 58 с.
7. Седжвик Р. Алгоритми на Java / Седжвик Р., Уэйн К. — 4-е изд. — М.: Вильямс, 2012. – 848 с.
8. Conrad S. Will Corpus Linguistics Revolutionize Grammar Teaching in the 21st Century? / S. Conrad // *Tesol Quartely* – 2000. – № 34. – P. 548-560.

Стаття надійшла до редакції 17.04.2014 р.

Мясникова В. В., магістрант,
Институт филологии КНУ имени Тараса Шевченка, г. Киев

АВТОМАТИЧЕСКАЯ ФОНЕТИЧЕСКАЯ ТРАНСКРИПЦИЯ И ТРАНСЛИТЕРАЦИЯ УКРАИНСКОГО ТЕКСТА

Статья посвящена комплексному рассмотрению программы автоматической фонетической транскрипции и транслитерации украинского текста «TranSScripT». Особое внимание обращается на принципы работы программы, которая обеспечивает автоматическую фонетическую транскрипцию, транслитерацию и обратную транслитерацию текстовых массивов. Подчеркивается уникальность исследования и его теоретическое и прикладное значение.

Ключевые слова: автоматическая обработка текста, алгоритм, транслитерация, фонетическая транскрипция.

V. Miasnikova, Student,
Institute of Philology Taras Shevchenko National University of Kyiv

THE AUTOMATIC PHONETIC TRANSCRIPTION AND TRANSLITERATION OF THE UKRAINIAN TEXT

The article is devoted to a complex review of the program of the automatic phonetic transcription and transliteration of the Ukrainian text «TranSScripT». Particular attention is paid to the principles of the program, which provides automatic phonetic transcription, transliteration and inverted transliteration of the text. The uniqueness of the research, its theoretical and practical value are emphasized.

Keywords: algorithm, automatic text processing, phonetic transcription, transliteration.

УДК 811.581'367.628

Нестеренко О. О., викл.,
Київський національний лінгвістичний університет

ОСОБЛИВОСТІ СИНОНІМІЧНОЇ ДИФЕРЕНЦІАЦІЇ ВИГУКІВ СУЧАСНОЇ КИТАЙСЬКОЇ МОВИ

У цій статті розкрито проблему синонімічних зв'язків вигуків сучасної китайської мови. Розглянуто питання синонімічної диференціації, особливостей побудови синонімічних рядів вигуківих одиниць, а також моделі формування синонімічних пар.

Ключові слова: вигук, сучасна китайська мова, синонімія, синонімічна модель.

Одиниці мовлення не існують ізольовано, а поєднуються відношеннями спорідненості та протиставлення, лексико-семантичними зв'язками, зокрема синонімічними, що утворюють цілісний комплекс.

Існує два підходи до визначення синонімів: власне семантичний та операційно-семантичний. Згідно з першим синоніми визначаються як слова, що мають одне й те саме лексичне значення, але відрізняються відтінками (В. Гумбольдт [8], О. О. Потєбня [8], А. О. Брагіна [3]). Прихильники другої концепції пропонують вважати синонімами лише такі слова, відмінності між якими можуть нейтралізуватися (нейтралізація семантичних відмінностей синонімів) (Ю. Д. Апресян [1], Д. Н. Шмельов [13]). Ряд цінних зауважень щодо синонімічних характеристик вигуків викладають О. О. Шахматов [12], В.В. Виноградов [4], А. І. Германович [5], хоча спеціально це питання не вивчалось.