УДК 004.91

АВТОМАТИЗОВАНИЙ ПЕРЕКЛАД З УКРАЇНСЬКОЇ МОВИ НА УКРАЇНСЬКУ ЖЕСТОВУ МОВУ: АЛГОРИТМІЧНА РЕАЛІЗАЦІЯ

Ю.В. Крак[1], О.В. Бармак[2], С.О. Романишин[2], А.В. Ізотов[1]

[1]Київський національний університет імені Тараса Шевченка, 01601, Київ, вул. Володимирська, 60
[2]Хмельницький національний університет МОН України, 29016, Хмельницький, вул. Інститутська, 11

Запропоновано алгоритмічну реалізацію інформаційної технології для перекладу з української мови на жестову мову. Побудовані інфологічні моделі словників української та жестової мов, пов'язані узагальнені граматичні конструкції для забезпечення автоматизованого перекладу. Наведені експериментальні результати для перевірки ефективності запропонованої інформаційної технології.

The article describes the algorithmic implementation of information technology for translation from Ukrainian into sign language. Infological model of Ukrainian and sign language dictionary, related generalized grammatical constructions for automatic translation were build. Analyzing the set of Ukrainian sentences and corresponding gesture sentences grammatical constructions for translation were built. Grammatical constructions contain parameters of word inflections for sentences. The experimental results to verify the effectiveness of the proposed information technology are given

Вступ та постановка задачі

Згідно зі статистикою Всесвітньої асоціації глухих, 4–10 % населення Європи страждають від проблем зі слухом. В Україні близько 300–500 тисяч людей з вадами слуху, функціонує 59 спецінтернатів для шкільного навчання дітей, 20 університетів навчають нечуючих студентів. Існує реальна потреба для людей з вадами слуху у створенні засобів рівноцінного спілкування у суспільстві. Важливою на сьогодні є проблема створення сучасних інформаційних технологій, що включають алгоритми, за допомогою яких можна було б створювати нові комп'ютерні системи навчання та комунікації для людей з вадами слуху. Для вирішення цієї проблеми запропонована концепція інформаційної технології невербального спілкування людей з вадами слуху [1]. Комплексна інформаційна технологія включає в себе функціональність по синтезу: рухів жестової мови, дактильної абетки, міміки промовляння на трьохмірній моделі людини та аналізу невербальних каналів комунікації.

Одним з напрямків концепції є створення автоматизованої системи перекладу з української на жестову мову. Жестова мова – це самостійна мова, яка виникла природно, та складається з комбінації жестів, кожен з яких виконується руками в поєднанні з мімікою, формою або рухом рота і губ, а також у поєднанні з положенням корпусу тіла. Жести умовно-схематичні, іноді придумуються на льоту, необов'язково мають візуальний зв'язок з позначуваним словом. Також, вони не є візуальною інтерпретацією звичайних мов; вони володіють своєю граматикою, можуть бути використані для обговорення самих різних тем: від простих і конкретних до піднесених або абстрактних [2].

Лексика розмовної жестової мови ще недостатньо вивчена, особливо ідіоматика, фразеологія, морфологія. Автоматичне приписування розмовної жестовій мові всіх форм словесної та письмової мови невірно. Не можна втиснути розмовну жестову мову глухих у традиційні моделі лінгвістики. Для опису фактів морфології розмовної жестової мови найбільше підходить опис за принципом: від значення – до форми. Значення «множинності», «часу», «приналежності», «закінчення дії» передаються у жестовому мовленні своїми способами. У залежності від значення змінюється і спосіб виконання жесту (якість руху).

Мета даної роботи – розробка інформаційної технології перекладу з української (УМ) на жестову мову (ЖМ).

Основна ідея, запропонована авторами, для реалізації інформаційної технології перекладу полягає у наступному[3]. При перекладі кваліфікованим перекладачем з однієї мови на іншу отримуються взаємно однозначні пари конструкцій, що передають сенс: речення на вхідній мові → відповідне речення на вихідній мові. Припускається, що дані пари конструкцій можна подати у вигляді певного узагальнення. Проаналізувавши певний об'єм отриманих при перекладі пар, можна, зафіксувавши порядок слідування слів у реченні, отримати узагальнення, у якому замість конкретних слів у реченні будуть використанні множини слів, які можуть використовуватись на цих зафіксованих місцях. Таким чином буде отримано досить невеликий (відносно загальної кількості речень) обсяг граматичних конструкцій перекладу.

Основні результати

Система автоматизованого перекладу з УМ на ЖМ передбачає створення моделі словників української та жестової мов з можливістю реалізації її для моделювання відповідних пов'язаних узагальнених граматичних конструкцій для цих мов. До розгляду беруться тільки прості речення. Вважається, що довільне складне речен-

ня можна подати як декомпозицію простих речень. Підмет і присудок у реченнях пов'язані предикативним зв'язком. Під простими реченнями розуміються речення з одним предикативним зв'язком. Як для УМ так і для ЖМ обмежимося такими типами простих речень: розповідні (стверджувальні, окличні та заперечні), питальні та спонукальні [3].

Для реалізації інформаційної технології автоматизованого перекладу запропоновані: інфологічні моделі: для словників УМ та ЖМ; узагальнені граматичні конструкції для перекладу; зв'язки між словами та жестами; алгоритм перекладу.

Інфологічна модель словників української та жестової мов. Для досягнення мети перекладу запропонована модель граматичного словника УМ [5, 6]. Модель побудована виходячи з флективності УМ. В основі запропонованої моделі УМ лежить теорія лексикографічних систем. Тобто, слова української мови подаються:

$$W = \{W_i : W_i = \{I_{i_1} \in I, F_{i_2} \in F, k, g \in G, s \in S, in \in In\}\}, \quad (1)$$

де W_i – параметри слова української мови ($i = 0, \cdots, N-1$, N – кількість слів у словнику); F – множина всіх можливих закінчень слів (флексій); k – номер позиції у слові-інфінітиві, з якої починається конкатенуватися флексія (можливі випадки коли $k = 0$ – для словоформи, яка повністю відміна від інфінітива); $G = \{g_1, \ldots, g_9\}$; $S = \{s_1, s_2\}$; $In = \{i_1, \ldots, i_{28}\}$ I – множина слів-інфінітивів української мови (для дієслів – інфінітиви, для іменників – слова у називному відмінку, однині тощо):

$$I = \{I_i : I_i = \{word\,\text{inf}, p \in P\}\}, \quad (2)$$

де $P \in \{p_1, \ldots, p_{30}\}$; $word\,\text{inf}$ – слово-інфінітив.

У роботах [5, 6] наведені формалізми для: 1) формування множин W, F, I ; 2) відтворення повного тексту слова за трьоми параметрам I_{i_1}, F_{i_2}, k , який реалізується відповідним оператором; 3) обернений оператор для визначення параметрів I_{i_1}, F_{i_2}, k для довільного слова.

Для жестової мови, в зв'язку з відсутністю у ній словозмін, структура словника буде дещо простішою. У невербальній комунікації мімічна складова відіграє дуже важливу, інколи ключову роль. Варто зазначити, що для синтаксису жестового мовлення для людей з вадами слуху характерна наявність немануального маркера: у запитальних реченнях, при загальному запитанні – піднятті брови, при окремому (приватному) запитанні – опущенні брови та нахилена вперед голова. При запереченні – негативні рухи головою, відповідний вираз обличчя. Для розповідних речень характерна наявність на обличчі емоційного забарвлення, що відповідає сенсу інформації, що передається. Множина жестів подається, як

$$Ges = \{Ges_i : Ges_i = \{word\,des, pges \in PGes, em \in Em\}\}, \quad (3)$$

де $word\,des$ – позначення жеста, $PGes \in \{p_1, \ldots, p_{11}, p_{13}, p_{22}\}$, $Em \in \{em_1, em_2, em_3\}$ – множина емоційних забарвлень жестів em_1. – розповідне емоційне забарвлення, em_2. – запитальне емоційне забарвлення, em_3 – інші емоційні забарвлення.

Інфологічну модель словників української та жестової мов подано наступним чином (рис.1):

Рис. 1. Інфологічна модель словників української та жестової мов

У таблиці «Словник української мови» міститься множина слів української мови (W), множина флексій (F) та інфінітивів (I) – в таблицях «Флексії» та «Інфінітиви» відповідно. Для кожного з слів зберігаються посилання на граматичні категорії, що визначають словозміну: «Множини», «Рід», «УКС» (узагальнена категорія словозміни). Множина жестів міститься в таблиці «Словник жестів» і містить посилання на слово, яким позначається даний жест і тематику, в якій він найчастіше використовується. Множини інфінітивів та жестів містять посилання на «Частини мови» для визначення приналежності їх до певних частин мови.

Для роботи з даною інфологічною моделлю треба запропонувати наступну алгоритмічну функціональність: додавання інфінітива, флексії та слова, отримання інфінітива слова, його флексії, визначення параметрів словозміни слова.

Додавання інфінітива проходить наступним чином:

- *Вхід*: слово-інфінітив та частина мови
- Визначаємо [№ частини мови] з таблиці «Частини мови» (4)
- Додаєм [інфінітив] і [№ частини мови] в таблицю «Частини мови»

Додавання слова та флексії:

- *Вхід*: слово української мови, інфінітив, множина, рід, узагальнена категорія словозміни
- Визначення [№ множини], [№ роду], [№ узагальненої категорії словозміни], [№ інфінітива] вибіркою з таблиць «Множини», «Рід», «УКС», «Інфінітиви»
- Визначаємо k = позиція слова, починаючи з якої воно відрізняється від інфінітива
- l = довжина слова
- флексія = RIGHT(слово,l-i)
- Якщо флексія не порожня (5)
- Якщо існує [флексія] в таблиці «Флексії» то визначаємо [№ флексії]
- Інакше додаєм флексію в таблицю «Флексії» та отримуємо її номер
- Кінець якщо
- Інакше визначаємо номер порожньої флексії з таблиці «Флексії»
- Кінець якщо
- Додаєм [№ інфінітива], [№ флексії], [позицію флексії], [№ множини], [№ роду], [№ узагальненої категорії словозміни] в таблицю «Словник української мови»

Отримання інфінітива, флексії, множини, роду, узагальненої категорії словозміни слова:

- *Вхід*: слово української мови
- Визначаєм [№ інфінітива] для слова з таблиці «Словник української мови» та [інфінітив] з таблиці «Інфінітиви» по зв'язаному полю [№ інфінітива]
- Визначаєм [№ флексії] для слова з таблиці «Словник української мови» та [флексію] з таблиці «Флексії» по зв'язаному полю [№ флексії]
- Визначаєм [№ множини] для слова з таблиці «Словник української мови» та [множину] з таблиці «Множини» по зв'язаному полю [№ множини] (6)
- Визначаєм [№ роду] для слова з таблиці «Словник української мови» та [рід] з таблиці «Рід» по зв'язаному полю [№ роду]
- Визначаєм [№ УКС] для слова з таблиці «Словник української мови» та [УКС] з таблиці «УКС» по зв'язаному полю [№ УКС]

Отримання частини мови для слова:

- *Вхід*: слово української мови
- *Визначаєм [№ інфінітива] для слова з таблиці «Словник української* мови» (7)
- Для ориманого [№ інфінітива] з таблиці «Інфінітиви» визначаєм [№ частини мови] та [частину мови] з таблиці «Частини мови» по зв'язаному полю [№ частини мови]

Інфологічна модель пов'язаних узагальнених граматичних конструкцій для перекладу. Після отримання інфологічного подення словників української та жестової мов потрібно побудувати пов'язані узагальнені граматичні конструкції речень УМ та ЖМ. Під реченням розумітимемо послідовність слів мови та знаків пунктуації. Під граматичною конструкцією речення розумітимемо послідовність слів мови, що відносяться до частин мови, які передають сенс. Тобто у речень відкидатимемо службові слова. Вважатимемо, що сенс передається наступними частинами мови: іменники (p_1), числівники порядкові (p_2), числівники кількісні (p_3), числівники (p_4), числівники типу «два» (p_5), дієслова недоконаного і доконаного виду (p_6), дієслова доконаного виду (p_7), дієслова недоконаного виду (p_8), дієприкметники (p_9), прикметники (p_{10}), займенники (p_{11}), прислівники (p_{13}), дієприслівники (p_{22}). Розрізнятимемо граматичні конструкції по кіль-

кості слів, що в них входять. Елементами граматичних конструкцій (на відміну від речень) будуть множини слів, що зустрічатимуться у реченнях, які складають навчальну вибірку. Отже:

$$Gr = \{Gr_i = \{word_j \in W \,||\, seq_j \in Seq \,||\, p_j \in P, s_j \in S, g_j \in G, i_j \in I, num\},$$
$$GStr_i = \{ges_j \in Ges \,||\, gseq \in Gseq \,||\, pges \in PGes, num, gesnum\}\}, \quad (8)$$

де num – номер елемента в структурі речення української мови, $gesnum$ – номер елемента в структурі речення жестової мови, $G = \{g_1,\ldots,g_9\}$; $S = \{s_1, s_2\}$; $In = \{i_1,\ldots,i_{28}\}$ – параметри словозміни,

$$Seq = \{Seq_i : Seq_i = \{word_j \in W : word_j \in W, n\}\},$$

де n – номер слова в послідовності,

$$Gseq = \{Gseq_i : Gseq_i = \{ges_j \in Ges : ges_j \in Ges, n\}\},$$

де n – номер слова в послідовності.

Після отримання множини граматичних конструкцій проведемо узагальнення кожної із множин елементів, які входять в неї. Узагальнення розумітимемо як подання кожної із отриманих множин як комбінацію конкретних граматичних категорій та частин мови:

$$GGr = \{GGr_i = \{PP_j, SS_j, GG_j, II_j, num\},$$
$$GStr_i = \{PPges, num, gesnum\}\}$$

$$PP = \{PP_i : PP_i = \{p_i \in P\}\}, \quad SS = \{SS_i : SS_i = \{s_i \in S\}\}, \quad II = \{II_j : II_j = \{i_j \in I\}\}, \quad (9)$$
$$GG = \{GG_i : GG_i = \{g_i \in G\}\}, \quad PPges = \{PPges_i : PPges_i = \{pges_i \in Pges\}\}$$

Елементами узагальнених граматичних конструкцій є множини параметрів словозміни, які можуть або не можуть використовуватись в фіксованій позиції речення (рис. 2). Крім того, до множин параметрів словозміни можуть додаватись або вилучатись послідовності слів.

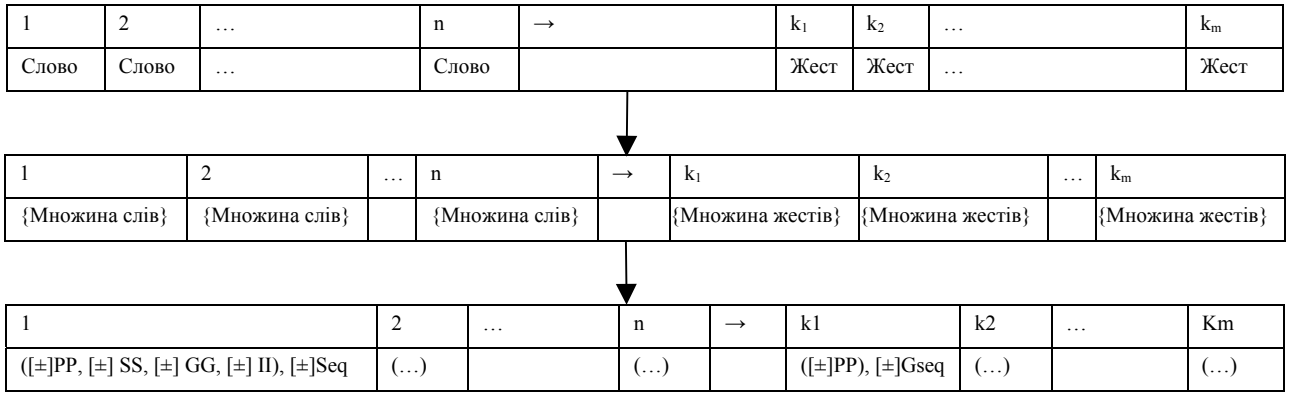

Рис. 2. Створення узагальнених граматичних конструкцій перекладу

Приклади узагальнених граматичних конструкцій перекладу наведено в таблиці:

Таблиця. Приклади елементів узагальнених граматичних конструкцій перекладу

((p1),(s1),(g1),(i1))	Числівники однина чоловічий рід називний відмінок
((p1,p2),(s1),(g1),(i1))	Числівники та іменники однина чоловічий рід називний відмінок
((p1),(s1),(g1),())	Числівники однина чоловічий рід будь-який відмінок

((p1),(s1),(g1),(-i2,-i3))	Числівники однина чоловічий рід будь-який відмінок крім родового та давального
((p1),(s1),(g1),(i1),+(він, вона))	Числівники однина чоловічий рід називний відмінок та множина слів {він, вона}
((p1),(s1),(g1),(i1),-(один, два))	Числівники однина чоловічий рід називний відмінок крім множини слів {один, два}

Для реалізації запропонованого підходу було сформовано множину речень, отриманих з програми-комплексу „Українська жестова мова"[7], який використовується в навчальних закладах для нечуючих з метою оволодіння жестовою мовою. Речення в моделі даних було об'єднано в структури, отримані шляхом узагальнення за формулою (8). Структури речень містять замість слів їх послідовності, кожна з яких може використовуватися при побудові речення. Послідовності можуть містити як окремі слова, так і різноманітні множини. Наприклад, речення «він іде», «вона іде», «час йде» було об'єднано в одну послідовність «{він, вона, час, хто, …} іде». Крім того, послідовності можуть містити категорії словозміни (p , s , g , i): частину мови (p), число (s), рід (g) та узагальнену категорію (i) (яка включає відмінок, особу, час, стан, спосіб), перетин яких визначає множину слів.

Для подання граматичних конструкцій (8) запропонована наступна інфологічна модель (рис. 3).

Рис. 3. Інфологічна модель для пов'язування структур речень УМ зі структурами речень ЖМ

Структури речень зберігаються в таблиці «Структури речень», структури жестових речень – в «Структури жестових речень». Дані таблиці містять інформацію про тип речення та тематику речення. «Елементи структур речень» можуть містити бути як окремі слова (№ слова), так і послідовності слів (№ послідовності). Для збереження порядку елементів структури речення кожен з елементів має порядковий номер. «Елементи структур жестових речень» також можуть містити жести та послідовності жестів. Для забезпечення відповідності при перекладі також в даній таблиці зберігається порядковий номер елемента структури речення українською мовою.

Для роботи з даною інфологічною моделлю потрібно запропонувати наступну алгоритмічну функціональність: створення послідовності слів, створення послідовності жестів, створення структури речення, створення структури жестового речення, отримання № структури речення для речення українською мовою, отримання номера структури жестового речення за номером структури речення українською мовою.

Створення послідовності слів:

- `На вхід подається послідовність слів`
- `Створюємо запис в таблиці «Послідовності слів», який містить перше слово послідовності та його порядковий номер` (10)
- `Отримуємо створений [№ послідовності]`
- `Для кожного з наступних слів послідовності додаємо [№ послідовності], [слово], № слова в послідовності] в таблицю «Послідовності слів»`

Створення послідовностей жестів аналогічне створенню послідовностей слів.

Створення структури речення:

- На вхід алгоритму подається структура речення, яка складається з слів або послідовностей слів, тип та тематика речення
- Створюємо запис в таблиці «Структури речень», який містить [тип речення] та [тему], отримуємо [№ структури речення] для створеного запису
- Для кожного елемента структури речення
- Якщо елемент – слово визначаємо [№ слова] з таблиці «Словник української мови»
- Додаємо запис в таблицю «Елементи структур речень», який містить [№ структури речення], [№ слова], [Знак пунктуації], [№ в структурі]
- Кінець якщо (11)
- Якщо елемент - послідовність
- виконуємо пошук її в таблиці «Послідовності слів»
- Якщо послідовність знайдена визначаємо [№ послідовності]
- Інакше створюємо нову послідовність за алгоритмом, описаним вище та визначаємо [№ послідовності]
- Додаємо запис в таблицю «Елементи структур речень», який містить [№ структури речення], [№ послідовності], [Знак пунктуації], [№ в структурі]
- Кінець якщо

Створення структури жестових речень аналогічне створенню структури речень, за винятком того, що в таблиці «Елементи структур жестових речень» міститься [№ структури в жестовому реченні] та [№ структури в реченні] для забезпечення відповідності між елементами речень УМ та ЖМ. Після створення структури жестового речення, відповідної структурі речення УМ в таблиці «Структури речень» вказується номер відповідної жестової конструкції.

Отримання номера структури для речення українською мовою:

- На вхід алгоритму подається речення українською мовою
- Для кожного слова речення знаходимо його № з таблиці «Словник української мови»
- Та формуємо індексовану послідовність номерів sentence
- Знаходимо список номерів структур речень list, для яких в таблиці «Елементи структур речень» даний sentence[1] міститься в послідовності слів або в полі [№ слова] та [№ в структурі] = 1 (12)
- Для всіх наступних слів речення серед структур з списку list шукаємо ті, які містять sentence[i] в послідовності слів або в полі [№ слова] та для яких [№ в структурі] = i
- Якщо така структура знайдена – повертаємо її номер

Шляхом аналізу отриманих множини структур речень було отримано множину узагальнених конструкцій української та жестової мов, які містять тільки категорії словозміни, та не містять окремих слів, або множин слів. Для кожного елемента структури речення було виділено множину категорії словозміни (p, s, g, i) та порядок даного елемента в конструкції (8).

Створення узагальнених конструкцій дозволило на основі аналізу категорій словозміни (p, s, g, i) слів у реченні та його порядкового номеру отримувати структуру жестового речення, що йому відповідає.

Інфологічна модель для узагальнених граматичних конструкцій відображена на рис.4.

Рис. 4. Інфологічна модель зв'язків між узагальненими структурами речень та узагальненими жестовими конструкціями

Таблиця «Узагальнені структури» містить узагальнені конструкції речень української мови. Основними елементами даної таблиці є категорії словозміни p, s, g, i та порядковий номер елемента в структурі речення. Таблиця «Узагальнені жестові структури» містить узагальнені конструкції жестових речень, основними елементами є частина мови та № елемента в структурі. Для роботи з даною інфологічною моделлю потрібно запропонувати наступну алгоритмічну функціональність: отримання № узагальненої структури для речення української мови, отримання узагальненої жестової конструкції за № узагальненої структури.

Для створення узагальнених жестових конструкцій запропоновано наступні алгоритмічну функціональність:

- `Для кожної структури struct з таблиці «Структури речень»`
- `Для кожної послідовності слів в структурі struct визначаємо параметри словозміни, спільні для її елементів` (13)
- `Шукаємо узагальнену структуру речення української мови в таблиці «Узагальнені структури», яка містить визначені параметри словозміни`
- `Якщо такої узагальненої структури не існує, додаємо узагальнену структуру GenStruct в таблицю «Узагальнені структури»`
- `Кінець якщо`
- `Для struct шукаємо відповідну їй структуру gesturestruct з таблиці «Структури жестових речень»`
- `Визначаємо параметри жестів, спільні для кожного з елементів структури gesturestruct та визначаємо узагальнену жестову конструкцію`
- `Якщо отримана узагальнена жестова конструкція вже існує, створюємо посилання на неї в GenStruct`
- `Інакше В таблицю «Узагальнені жестові конструкції» додаємо узагальнену жестову конструкцію та посилання на неї в GenStruct`
- `Кінець якщо`

Крім визначення структури речення необхідною для перекладу є наявність відповідності слів жестам. Використовуючи відповідності між структурами речень та структурами жестових речень було виділено множину відповідностей «слово → жест». В таблиці «Відповідність слів жестам» міститься № слова та № відповідних йому жестів.

Для отримання відповідностей «слово → жест» для кожної структури речень з таблиці «Структури речень» та відповідної їй жестової структури речень з таблиці «Структури жестових речень» було отримано відповідності між послідовностями слів та послідовностями жестів. Використовуючи дані відповідності з таблиць «Послідовності слів» та «Послідовності жестів» було сформовано таблицю «Відповідність слів жестам». У випадку наявності не однозначних відповідностей «слово → жест» така інфологічна модель не дає можливості отримати чітке відношення між множиною слів та жестів. Якщо одному слову відповідає декілька жестів, для обрання правильного жесту використовуються відношення між реченнями «Структури речень» та «Структури жестових речень» з врахуванням тематики речення, що перекладається. У випадку, якщо кількома словам відповідає один жест труднощів в отриманні відповідності «слово → жест» не виникає. Крім то-

го, для таких жестів можна отримати список слів-синонімів, які їм відповідають. Для заповнення таблиці «Відповідність слів жестам» відбувалось за таким алгоритмом:

- Для кожної структури речення з «Узагальнені структури» визначаємо відповідну їй структуру з «Структури жестових речень»
- Кожній послідовності слів в структурі речення ставимо у відповідність послідовність жестів (14)
- Для кожного слова послідовності визначаємо його номер та номер жесту з відповідної жестової послідовності
- Номер слова та номери відповідних жестів додаємо в таблицю «Відповідність слів жестам»

Для автоматизованого перекладу тексту з української мови на жестову мову запропоновано наступний алгоритм (рис. 5):

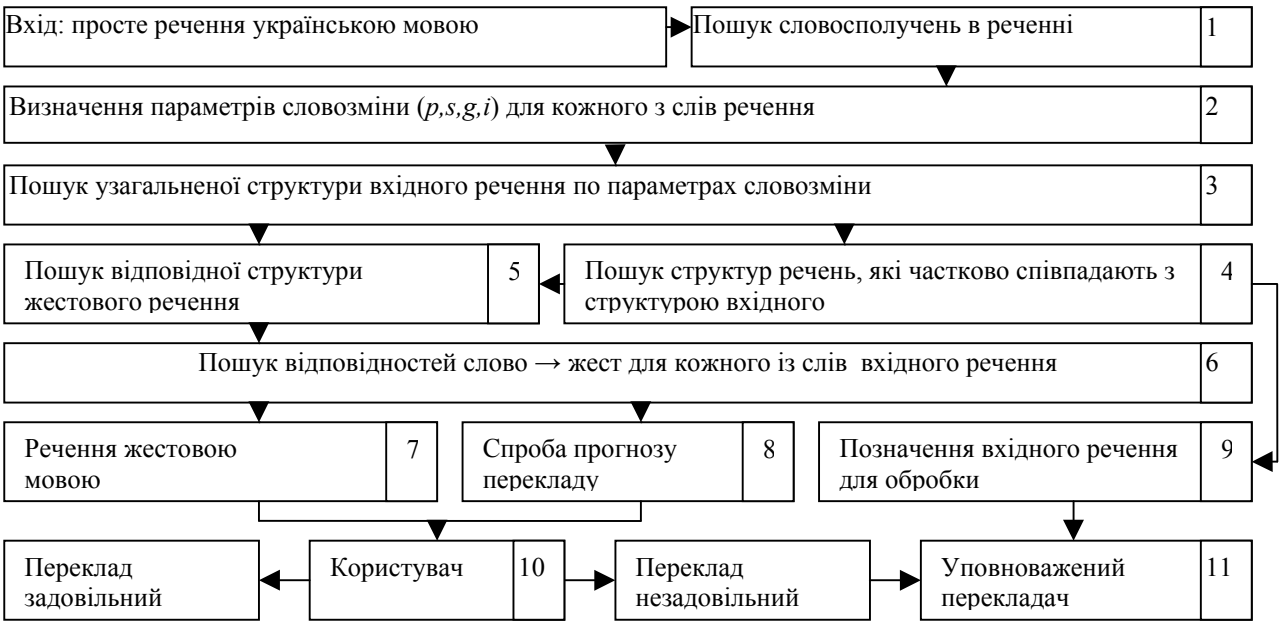

Рис.5. Алгоритмічна схема автоматизованого перекладу

1. Речення поступає на вхід, виконується пошук словосполучень в реченні:

- *Вхід*: речення українською мовою
- Речення подаємо у вигляді індексованої послідовності слів list[] = множина речень слова
- list2[] = null
- k = кількість слів в реченні, i=0, j=0
- Повторювати поки i<k-1
- Якщо list[i] + ' ' + list[i+1] існує в таблиці «Словник української мови» (15)
- list2[j] = list[i] + list[i+1]
- i = i+2
- j = j+1
- Інакше i = i+1
- Кінець циклу
- Отриманий список list2 містить індексовану послідовність слів та словосполучень

2. Використовуючи Словник української мови, для кожного з слів та словосполучень з списку *list2* з таблиці «Словник української мови», використовуючи вище описані алгоритми визначаються параметри словозміни p, s, g, i.

3. На основі параметрів словозміни кожного з слів визначається узагальнена конструкція української мови:

- *Вхід*: параметри словозміни та номер по порядку слів речення
- В таблиці «Узагальнені структури» виконуємо пошук узагальненої структури речення, яка відповідає вхідним параметрам: перший елемент структури містить категорії словозміни, відповідні першому слову речення, другий елемнт – другому слову, для кожного з слів речення (16)

4. Якщо відповідна узагальнена конструкція не знайдена, виконуємо пошук схожий за структурою конструкцій:

- *Вхід*: параметри словозміни та номер по порядку слів речення
- В таблиці «Узагальнені структури» виконуємо пошук узагальненої структури речення, яка частково відповідає вхідним параметрам: всі елементи структури, крім одного відповідають словам вхідного речення (17)

5. Знайшовши узагальнену структуру речення, отримуємо відповідну йому узагальнену жестову конструкцію з таблиці «Узагальнені жестові конструкції по зв'язаному полю [№ узагальненої жестової конструкції].

6. Для кожного з слів вхідного речення виконується пошук відповідностей «слово → жест», використовуючи таблиці «Відповідність слів жестам», «Словник української мови» та словник жестів.

7. Якщо знайдено узагальнену структуру речення української мови, відповідну їй узагальнену жестову конструкцію та всі відповідності «слово → жест», то для отриманої конструкції жестової мови, використовуючи відповідності «слово → жест» виводиться результат перекладу.

8. Якщо знайдено узагальнену структуру речення української мови, яка частково співпадає з структурою вхідного речення, виконується спроба прогнозу перекладу. Замість узагальненої структури речення використовується структура, яка частково співпадає з вхідною та на її основі, використовуючи відповідності «слово → жест» будується жестове речення.

9. Якщо не знайдено узагальнених структур, які частково співпадають з структурою вхідного речення, то воно додається в статистичну базу для подальшої обробки. Якщо для якогось з слів речення не знайдено відповідного жесту, слово позначається в статистичній базі для подальшої обробки.

10. Отримавши результат, користувач вибирає один з варіантів:

– Результат задовільний. Речення позначається як правильне і добавляється в статистичну базу.

– Результат не задовільний і правиться вручну. Після цього в статистичну базу вноситься і виправлений варіант і запропонований варіант, для забезпечення прийняття рішень надалі.

11. У випадку, якщо не знайдено узагальнену структуру речення, відповідності «слово → жест», або результат перекладу не задовільний - уповноважений перекладач може додати нову конструкцію перекладу, змінити уже існуючу, встановити відповідність «слово → жест» або додати новий жест.

Реалізація побудови моделі словників української та жестової мов. Для заповнення інфологічної моделі даними по алгоритмах (1–2) було розроблено застосування, яке на вхід якого подається слово української мови. За допомогою (3–5) слово розбивається на комбінацію інфінітива та флексії та записується в базу даних. Потрібна для цього процесу функціональність реалізована у класі AllwordsBuilder (рис. 6).

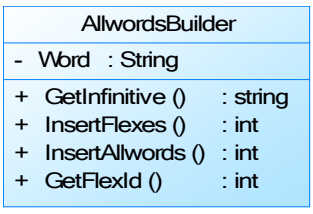

Рис.6. Діаграма класу AllwordsBuilder

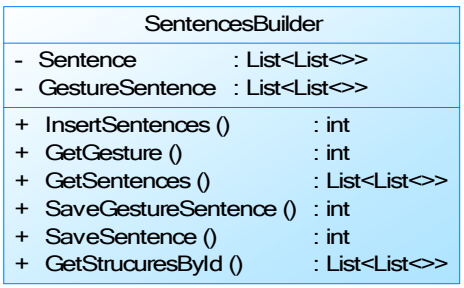

Рис.7. Діаграма класу SentencesBuilder

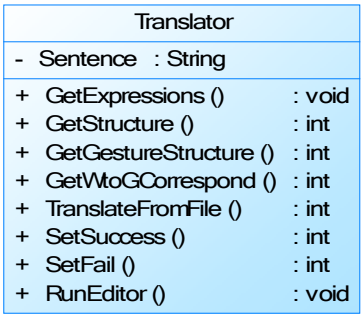

Рис.8. Діаграма класу Translator

Основним полем класу AllwordsBuilder є Word, яке містить вхідне слово українсько мовою. Основними методами є InsertAllwords() – у випадку, якщо Word – інфінітив воно заноситься в базу даних в таблицю «Інфінітиви». В іншому випадку виконується функція GetInfinitive() для отримання інфінітива даного слова, після чого для слова знаходиться спільна з інфінітивом частина та його флексія. Виконується функція GetFlexId() для отримання порядкового номера флексії з бази даних. Якщо такого номера не існує, то з допомогою функції InsertFlexes() додається нова флексія, після чого слово зберігається в таблиці «Словник української мови», як № інфінітива, № флексії, позиція флексії та параметри словозміни слова.

Реалізація побудовника узагальнених граматичних конструкцій. Для заповнення структур речень української мови за допомогою алгоритмів (6–8,10,11) розроблено клас SentencesBuilder (рис. 7). Основними полями класу є Sentence та GestureSentence. Основним методом класу є InsertSentences(), на вхід якого подавалось речення Sentence. Виконувався пошук в базі даних структур, які частково співпадають з вхідним реченням за алгоритмом (12). У випадку, якщо такі структури знайдено, слова вхідного речення додавались до даної структури речення (10). Якщо схожих структур не було, створювалась нова структура (11).

Для створення жестових конструкцій було розроблено відповідне застосування (рис. 9). Використовуючи методи GetSentences() за алгоритмом (12) виводилась множина структур речень відповідно до обраної теми.

При виборі структури речення з допомогою методу GetStrucuresById() виконувався пошук відповідної структури жестового речення. Для кожної з структур було побудовано відповідну їй структуру жестового речення.

Для додавання жестів в структуру жестового речення використовувався метод GetGesture(), який по позначенню жесту виконував пошук в базі даних. Після створення або редагування структури жестових речень та речень української мови зміни в структурах зберігались з допомогою методів SaveSentence() та SaveGestureSentence().

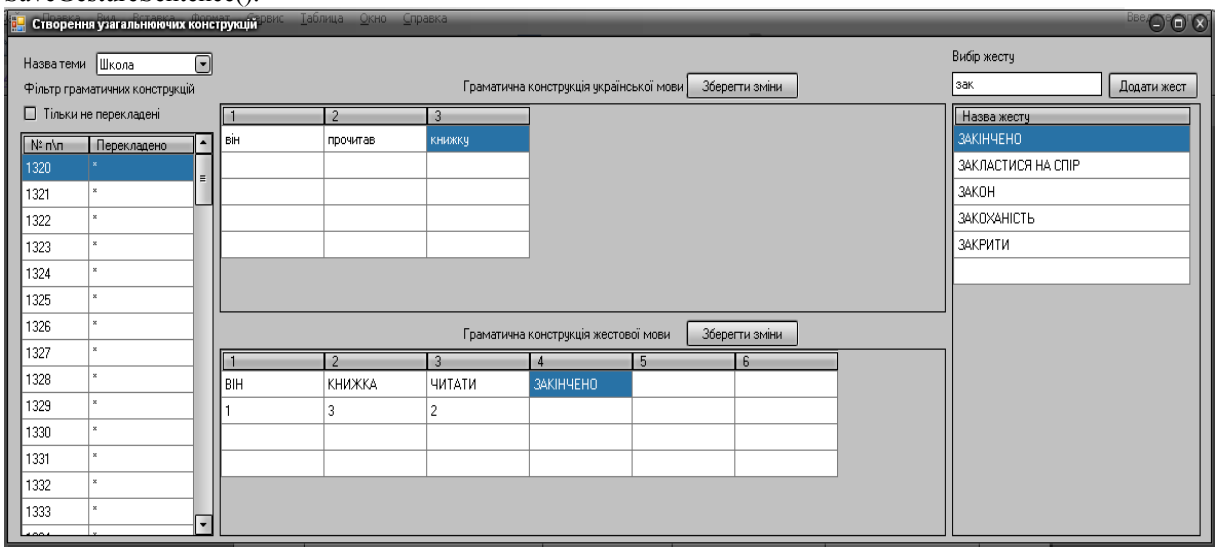

Рис. 9. Застосування для створення узагальнених жестових конструкцій

Реалізація інформаційної системи для автоматизованого перекладу. Для перевірки ефективності запропонованої технології перекладу було реалізовано інформаційну систему автоматизованого перекладу у вигляді застосування (рис. 10).

Для реалізації застосування було побудовано клас Translator (рис.8). Основним полем класу є Sentence, в якому зберігається вхідне речення українською мовою. Отримавши речення для перекладу за алгоритмом (15) використовуючи метод GetExpressions() виконується пошук словосполучень в вхідному реченні. Після знаходження словосполучень використовуючи методи GetStructure() та GetGestureStructure() за алгоритмами (16–17) визначається узагальнена структура речення та відповідна їй узагальнена жестова конструкція. Використовуючи метод GetWtoGCorrespond() визначаються відповідності «слово → жест» для кожного з слів вхідного речення, після чого виводиться результат перекладу. Отримавши переклад, можна позначити його як задовільний (метод SetSuccess()), незадовільний (метод SetFail()), або редагувати його (метод RunEditor()).

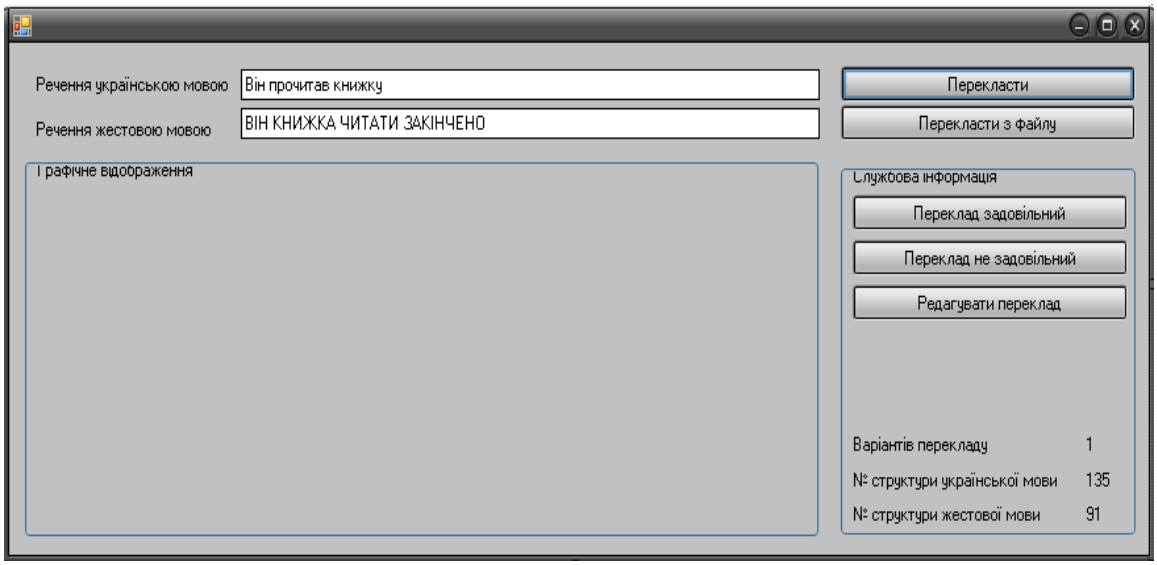

Рис. 10. Система автоматизованого перекладу

Експериментальні результати. За допомогою запропонованої технології, було створено словник української та жестової мов, отримано множину з 2000000 слів української мови та 3200 жестів. В результаті перекладу множини речень з української на жестову мову та групування їх в граматичні конструкції було отримано 1050 конструкцій перекладу. Після проведення аналізу конструкцій перекладу було отримано 293 узагальнених конструкції перекладу. При тестуванні на реченнях, взятих з програми вивчення української жестової мови отримано автоматизований переклад без спотворення сенсу в 100 % випадків.

При тестуванні системи перекладу на множині простих 500 речень, взятих з газети «Наше життя» УТОГ, було отримано наступні результати: автоматичний переклад для 64 % речень; визначення граматичної конструкції перекладу для речення, відсутність при перекладі відповідностей слово-жест для 24 % речень; відсутність граматичних конструкцій перекладу для 12 % речень (з них для 55 % речень знайдено подібні жестові конструкції та запропоновано прогноз перекладу). Задовільний автоматизований переклад (автоматичний переклад та переклад з використанням правильних подібних конструкцій перекладу) за умови наявності відповідності слово-жест склав 94 %.

Висновки

Запропоновано інформаційну технологію автоматизованого перекладу з української на жестову мову та виконано її експериментальну перевірку. Створено інфологічне подання словників української мови, як множин інфінітивів, флексій, та параметрів словозміни слів та жестової мови як множину жестів та параметрів, що їх характеризують; розроблено механізм подання відношень між словами української та жестами жестової мов у вигляді узагальнених граматичних конструкцій та зв'язків між жестами і словами; розроблено інформаційну технологію для здійснення автоматизованого перекладу за допомогою отриманих моделей; алгоритмічно та програмно реалізовано інформаційну технологію для підтвердження її життєздатності.

Під час тестування технології отримано автоматичний переклад вхідних речень у 64 % випадках. При доповненні системи перекладу відповідностями слово-жест задовільний автоматизований переклад склав 94 %.

Подальші дослідження спрямовані на створення на основі запропонованої інформаційної технології веб-застосування, яке б дозволило виконувати переклад в режимі online.

1. *Кривонос Ю.Г., Крак Ю.В., Бармак О.В. та ін.* Інформаційна технологія для моделювання української мови жестів // Штучний інтелект. – 2009. – № 3 – С. 186–197.
2. *Зайцева Г.Л.* Жестовая речь. Дактилология: Учеб. для студ. высш. учеб. заведений. – М.: Гуманит. изд. центр ВЛАДОС, 2000. – С. 61.
3. *Крак Ю.В., Бармак А.В., Романишин С.А.* Метод обобщенных грамматических конструкций для автоматизированного перевода текстовой информации в жестовые аналоги // Кибернетика и системный анализ. – 2014. – № 1. – С. 130–138.
4. *Шульжук К.Ф.* Синтаксис української мови. – К. : Академія, 2004. – С. 152–160.
5. *Крак Ю.В., Бармак О.В., Романишин С.О.* Узагальнені граматичні конструкції для автоматизованого перекладу з української мови на українську жестову мову // Штучний інтелект – 2011. – № 3. – С. 136–146.
6. *Бармак О.В., Романишин С.О.* Комп'ютерна технологія для перекладу з української на жестову мову // Вісник Тернопільського національного технічного університету. – 2012. – № 1(65). – С. 195–208.
7. *Програма*-комплекс "УКРАЇНСЬКА ЖЕСТОВА МОВА" // http://www.mon.gov.ua/education/average/programs_gluh