

Міністерство освіти і науки України  
Національний університет “Львівська політехніка”

Лозинська Ольга Володимирівна



УДК 004.912+004.415

**СИСТЕМА КОМП'ЮТЕРНОГО ПЕРЕКЛАДУ УКРАЇНСЬКОЇ  
ЖЕСТОВОЇ МОВИ З ВИКОРИСТАННЯМ ГРАМАТИЧНО  
ДОПОВНЕНОЇ ОНТОЛОГІЇ**

10.02.21 – структурна, прикладна та математична лінгвістика

Автореферат  
дисертації на здобуття наукового ступеня  
кандидата технічних наук

Львів – 2016

Дисертацією є рукопис.

Робота виконана у Національному університеті «Львівська політехніка» Міністерства освіти і науки України.

Науковий керівник: доктор технічних наук, професор  
**Пасічник Володимир Володимирович**,  
Національний університет «Львівська політехніка»,  
професор кафедри інформаційних систем та мереж

Офіційні опоненти: доктор технічних наук, професор  
**Лупенко Сергій Анатолійович**,  
Тернопільський національний технічний університет  
імені Івана Пулюя,  
професор кафедри комп'ютерних систем та мереж  
  
кандидат технічних наук, старший науковий співробітник  
**Остапова Ірина Вадимівна**,  
Український мовно-інформаційний фонд НАН України

Захист відбудеться «11» березня 2016 р. о 16.00 годині на засіданні спеціалізованої вченої ради Д 35.052.05 у Національному університеті «Львівська політехніка» (79013, м. Львів, вул. С. Бандери, 12).

З дисертацією можна ознайомитися у бібліотеці Національного університету «Львівська політехніка» (79013, м. Львів, вул. Професорська, 1).

Автореферат розіслано «10» лютого 2016 р.

Учений секретар спеціалізованої вченої ради,  
доктор технічних наук, професор



Р.А. Бунь

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** Розвиток сучасної науки, комп'ютеризація суспільства, використання мультимедійних та інтернет-технологій створили умови для розроблення комп'ютеризованих систем комунікування людей із особливими потребами, зокрема людей з вадами слуху. Людям, які спілкуються жестовою мовою, необхідно надати комфортний доступ до сучасних інформаційних ресурсів, поданих в інформаційних системах та мережах. В Україні проживає близько 400 тисяч людей з вадами слуху. Для такої категорії громадян функціонує 59 спеціалізованих шкіл, у 20 українських університетах (зокрема Національний технічний університет України «Київський політехнічний інститут») готують фахівців. В освітньо-науковому та технологічному контекстах актуальним завданням є розроблення програмних засобів, які допомагають у вивченні та перекладі української жестової мови.

Через великий обсяг інформації, яка доступна нині користувачам комп'ютерів, неможливо забезпечити переклад усіх текстових і мультимедійних ресурсів жестовою мовою. Одним з підходів до вирішення цієї проблеми є розроблення комп'ютеризованої системи перекладу української словесної мови на українську жестову мову та навпаки.

Жестова мова (ЖМ) – це комунікативна система, висловлювання якої побудовані не на звуковій, а на жестикуляційно-мімічній основі (жести рук, артикуляція губ, міміка та емоції). Українська жестова мова (УЖМ) є природною візуально-просторовою мовою, яка використовується спільнотою тих, хто не чує. Ця мова незалежна від української словесної мови (УСМ), і не слід її плутати з калькованою жестовою мовою (КЖМ), граматика якої повністю відповідає УСМ.

Для вирішення проблеми доступу до інформаційних ресурсів осіб з вадами слуху розроблено словники та відеословники УЖМ, а також тренажери УЖМ. Проте немає ефективних засобів перекладу.

Переклад жестової мови на словесну можна поділити на два етапи – етап розпізнавання й анотування жестів та етап переклад анотованих жестів. Проблему розпізнавання й анотування жестів досліджували вчені Т. Гіп (Т. Hear) і Д. Гогг (D. Hogg), Дж. Ліхтенавер (J. F. Lichtenauer) та Г. Гольт (G. A. ten Holt), К. Маєрс (C. S. Myers) та Л. Ребінер (L. R. Rabiner), М. Загеді (M. Zahedi) та ін., для її вирішення розроблено спеціальні апаратно-програмні комплекси, такі як Microsoft Kinect, Acer Aspire One, Intel IntelliSense тощо. З іншого боку, проблема автоматичного перекладу анотованої української жестової мови ще не вирішена на належному рівні.

У наукових роботах Д. Стайна (D. Stein), Ф. Дрю (P. Dreuw), Г. Ней (H. Ney) досліджено та розроблено статистичні методи перекладу для англійської жестової мови. С. Моррісей (S. Morrissey) та Е. Вей (A. Way) для перекладу англійського тексту на ірландську жестову мову застосували методи перекладу на основі даних. У роботі іспанських вчених Р. Сан-Сегундо (R. San-Segundo), А. Переза (A. Pérez), Д. Ортіза (D. Ortiz), Л. Д. Д'Гаро (L. F. D'Haro), М. І. Тореза (M. I. Torres), Ф. Касакуберта (F. Casacuberta) використано підхід на основі правил для перекладу іспанської розмовної мови на іспанську жестову мову. Вчені

Я. Банджерот (J. Bungeroth) та Г. Ней запропонували статистичний підхід для перекладу німецької жестової мови. Для польської жестової мови вчені Н. Сужанська (N. Suszczanska), П. Жмаль (P. Szmal) та С. Куликов (S. Kulikow) розробили систему перекладу на основі правил Thetos, яка є доповнена такими процесами, як синтаксичний та семантичний аналізи вхідного тексту. Для перекладу іноземних жестових мов розроблені такі засоби: програма iCommunicator (для англійської жестової мови), програма ViSiCAST (для англійської жестової мови), програма TEAM Project (для англійської жестової мови), програма Artemis (для французької жестової мови). Проте ці засоби не підходять для перекладу УЖМ. Автору не відомі роботи, де б досліджувались застосовність наведених методів для перекладу УЖМ.

Українські вчені Ю. В. Крак, О. В. Бармак та ін. з Інституту кібернетики ім. В. М. Глушкова НАН України запропонували реалізувати інформаційну технологію автоматизованого перекладу з флективних мов на жестову з використанням інфологічних моделей словників. Проте для досягнення якісного комп'ютерного перекладу української жестової мови важливе застосування методів семантичного аналізу. Таким методом у дисертаційній роботі вибрано метод на основі граматично доповненої онтології (ГДО).

Дисертаційне дослідження присвячено розробленню системи перекладу анотованої української жестової мови на основі граматично доповненої онтології, яка не має аналогів у відомих наукових дослідженнях. Створені упродовж дослідження методи та програмні засоби опрацювання та перекладу української жестової мови можна використати під час розроблення програм перекладу з однієї мови на іншу, що полегшить спілкування осіб із вадами слуху між собою і з тими, хто не володіє жестовою мовою.

**Зв'язок теми дисертації з науковими програмами, планами, темами.** Подані у дисертації дослідження виконані відповідно до напрямів наукової діяльності кафедри інформаційних систем та мереж Національного університету «Львівська політехніка» та теми Міністерства освіти і науки України «Математичне та програмно-алгоритмічне моделювання процесів трансляції жестів у текст для спеціалізованих комп'ютерних систем» у 2011–2012 рр. (номер державної реєстрації 0111U001222). Автор розробила граматично доповнену онтологію для опису та перекладу анотованої української жестової мови, класифікувала правила перекладу для української жестової мови, створила систему комп'ютерного перекладу з української словесної мови на анотовану українську жестову мову та навпаки.

**Мета і задачі дослідження.** Метою дисертаційної роботи є розроблення системи комп'ютерного перекладу анотованої української жестової мови на основі граматично доповненої онтології.

Для досягнення поставленої мети розв'язано такі задачі:

- проаналізовано граматичні особливості української жестової мови та сучасні підходи до створення систем комп'ютерного перекладу жестових мов;
- виконано декомпозицію проблеми перекладу української жестової мови та спроектовано архітектуру програмно-алгоритмічного комплексу для її вирішення;

- побудовано систему правил для комп'ютерного перекладу анотованої української жестової мови з використанням граматично доповненої онтології;
- розроблено методи комп'ютерного перекладу анотованої української жестової мови з використанням граматично доповненої онтології;
- проведено експериментальні дослідження та виконано порівняльний аналіз отриманих результатів.

*Об'єктом дослідження є процеси машинного перекладу анотованої жестової мови.*

*Предметом дослідження є методи та засоби машинного перекладу анотованої української жестової мови з використанням граматично доповненої онтології.*

**Методи дослідження.** Для підготовки вхідних даних та інтерпретації отриманих результатів використано основні положення та методи математичної лінгвістики, методи на основі онтологій, статистичного аналізу, класифікації; для синтаксичного аналізу речень української словесної мови та української жестової мови розроблено метод семантично-синтаксичного розбору речень з використанням ГДО, який використовує граматику складових; для перекладу речень УЖМ у речення УСМ та навпаки використано метод трансформації дерева складових у дерево залежностей; під час розроблення моделі онтологічного словника та словника концептів «УСМ – УЖМ» використано апарат теорії реляційних баз даних; для розроблення програмного забезпечення застосовано методи об'єктно-орієнтованого програмування; для машинного перекладу УЖМ використано відомі статистичні методи перекладу, методи перекладу на основі правил.

**Наукова новизна одержаних результатів.** У дисертаційній роботі вирішено актуальне наукове завдання комп'ютерного перекладу української жестової мови. Водночас:

*вперше розроблено*

- систему правил перекладу з української словесної мови на анотовану українську жестову мову з використанням граматично доповненої онтології, що дало змогу підвищити відсоток правильно перекладених речень з 57,2 % до 93,2 %;
- правила семантично-синтаксичного розбору речень для української словесної та анотованої української жестової мов, що дало змогу здійснити семантичний аналіз речень обома мовами;
- архітектуру програмно-алгоритмічного комплексу перекладу анотованої української жестової мови, яка відрізняється від відомих іноземних аналогів додаванням модуля перекладу на основі граматично доповненої онтології, що дало змогу автоматично перекладати УЖМ;
- словник концептів «Українська словесна мова – Українська жестова мова» для перекладу анотованої української жестової мови, що дало змогу зменшити відсоток неправильно перекладених тестових речень на 8,25 %;

*одержав подальший розвиток*

- метод наповнення онтологій з використанням предметно-орієнтованих мов, що дало змогу розробити граматично доповнену онтологію.

**Практичне значення одержаних результатів.** Усі дослідження виконано з метою практичної реалізації. Результати досліджень стали основою програмно-алгоритмічного комплексу перекладу анотованої української жестової мови.

Результати дисертаційних досліджень використано для розроблення системи комп'ютерного перекладу анотованої української жестової мови та впроваджено у Львівській спеціальній загальноосвітній школі-інтернаті Марії Покрови для глухих дітей.

**Особистий внесок здобувача.** Усі теоретичні та експериментальні дослідження, розроблення методів, алгоритмічного та програмного забезпечення виконала автор самостійно. У роботах, опублікованих у співавторстві, здобувачеві належить: розроблення граматично доповненої онтології, яка використана для семантично-синтаксичного розбору речень української словесної мови та української жестової мови, обґрунтування доцільності використання граматично доповненої онтології для перекладу української жестової мови [1, 3]; аналіз відомих онтологічних словників та систем побудови онтологій, наповнення граматично доповненої онтології, проведення експериментів [2, 4]; розроблення словника концептів української жестової мови та його використання для перекладу УЖМ, проведення експериментів [5, 15]; розроблення основних правил обходу дерев синтаксичного розбору речень української словесної мови [6, 13, 14]; застосування моделі статистичного машинного перекладу для української жестової мови та проведення експериментів [8]; дослідження лінгвістичних особливостей української жестової мови на прикладі простих речень та створення корпусу паралельних текстів «Українська словесна мова – Українська жестова мова» [9, 16, 17]; адаптація відомих інформаційних технологій перекладу до української жестової мови [10, 12]; обґрунтування доцільності використання інформаційних технологій для розроблення комп'ютеризованих систем комунікації людей із особливими потребами [19, 20].

**Апробація результатів дисертації.** Наукові та практичні результати роботи доповідалися та обговорювалися на: Міжнародній науково-технічній конференції «Комп'ютерні науки та інформаційні технології» (м. Львів, 2012); Міжнародній науково-технічній конференції «Перспективні технології і методи проектування МЕМС» (с. Поляна, 2013, м. Львів, 2014); Міжнародній науково-практичній конференції «Інформаційні управляючі системи та технології» (ІУСТ-2013) (м. Одеса, 2013); I Міжнародній науково-практичній конференції «Співпраця між університетами та підприємствами в галузі IT-аутсорсингу (ISSUBITO 2013)» (м. Харків, 2013); Міжнародній конференції «Наука как основа мирного диалога» (м. Донецьк, 2014); V Міжнародній науково-технічній конференції «Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління» (м. Полтава, 2014); IV Міжнародній науково-практичній конференції «Математика. Інформаційні технології. Освіта» (м. Луцьк, 2015).

Матеріали досліджень апробовані на чотирьох кафедральних наукових семінарах у повному обсязі.

У 2013 р. робота під назвою «Освітні інформаційні технології для глухих та слабкочуючих. Напрацювання та досвід Національного університету «Львівська політехніка» (автори: М. В. Давидов, О. В. Лозинська, О. В. Вербич) експонувалась на V Національній виставці «Інноватика в сучасній освіті», м. Київ.

**Публікації.** Результати дисертації опубліковано у 20 наукових працях, з яких 7 статей у фахових наукових виданнях, 3 публікації у закордонних виданнях (Польща, Росія), 1 публікація у виданні, що не належить до фахових наукових видань України, 8 публікацій у матеріалах і тезах конференцій, 1 наукове видання-брошура (препринт).

**Структура та обсяг дисертації.** Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, додатків та списку використаних джерел з 119 найменувань. Загальний обсяг дисертації 160 сторінок. Основний зміст викладено на 132 сторінках друкованого тексту. Робота містить 46 рисунків та 30 таблиць. У додатках наведено опис наборів тестових даних та акти впровадження.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі до дисертації обґрунтовано актуальність досліджень, сформульовано мету та її завдання. Наведено характеристику наукової новизни та практичної цінності отриманих результатів. Показано зв'язок роботи з науковими програмами, планами, темами.

У **першому розділі** розглянуто жестову мову як об'єкт комп'ютерного перекладу. Описано особливості жестової мови порівняно зі словесною мовою. Розглянуто відомі системи перекладу жестової мови, порівняно методи запису жестової мови, проаналізовано організацію словників жестової мови та онтологічних словників.

Проведено огляд сучасних досліджень, які виконуються з метою перекладу іноземних жестових мов. Для перекладу на жестову мову відомі різні методи, зокрема методи перекладу на основі прикладів, методи перекладу на основі правил, методи перекладу на основі даних та статистичні методи перекладу. Керівниками досліджень є Р. Сан-Сегундо та С. Балдассаррі (Іспанія), проф. Г. Ней, Д. Стайн та Я. Банджерот (Німеччина), Т. Дасгупта (Індія), Я. Каніс та Л. Мюллер (Чехія), проф. Л. Зійл (Стелленбос).

Під час дослідження встановлено, що переклад анотованої української жестової мови є складною проблемою, вирішення якої потребує розв'язання низки задач, пов'язаних з усуненням неоднозначності слів, синтаксичним розбором жестових та словесних речень, збереженням та використанням знань експерта-перекладача.

Незважаючи на збільшення кількості наукових робіт, пов'язаних із синтаксичними особливостями української жестової мови, досі немає відкритих для використання програм автоматичного синтаксичного аналізу речень українською жестовою мовою.

Огляд відомих методів перекладу жестових мов показав, що найбільше застосовують для перекладу анотованої української жестової мови метод машинного перекладу на основі правил. Складність розроблення системи автоматичного перекладу анотованої української жестової мови посилюється відсутністю великих словників та корпусів УЖМ. Оскільки жестові мови використовують для подання інформації різні жести і вирази замість звуків,

адаптування систем перекладу словесних мов для перекладу жестової мови потребує розроблення системи запису всіх елементів жестової мови, створення відповідних перекладацьких словників і системи граматичних правил для розбору жестової мови. Крім того, значна лінгвістична відмінність жестової мови від словесної ускладнює процес перекладу.

Під час дослідження встановлено, що тривалий час джерелом знань для систем перекладу були тлумачні та двомовні словники. Через те, що тлумачні словники не володіють необхідною повнотою і структурою для вираження семантичних зв'язків між словами і поняттями, замість них почали використовувати онтологічні словники. Сучасні онтологічні словники містять не лише конкретні значення слів, а й лексичні (антонімія, слова-відношення, номіналізація та ін.) та семантичні (гіперонімія/гіпонімія, меронімія/голонімія та ін.) зв'язки між ними, що дає змогу використовувати їх для усунення багатозначності слів на основі цих зв'язків. До таких онтологічних словників належать WordNet та ConceptNet для англійської мови, plWordNet для польської мови, CWN для китайської мови, WOLF для французької мови, MultiWordNet для італійської мови, BalkaNet project для шести європейських мов (болгарської, чеської, грецької, румунської, турецької та сербської), GermaNet для німецької мови, IndoWordNet для індійської мови, проект RussNet для російської мови.

Визначено, що ключові завдання під час створення системи перекладу анотованої української жестової мови – розроблення методів та засобів для ефективного перекладу УЖМ, оптимізація наповнення граматично доповненої онтології, розроблення системи правил перетворення української словесної мови у жестову та навпаки. На основі проведеного аналізу детально сформульовано завдання дослідження.

У **другому розділі** розроблено інфологічну модель словника концептів «Українська словесна мова – Українська жестова мова», розроблено математичну модель **граматично доповненої онтології**. Ця модель використана для семантично-синтаксичного розбору речень корпусу паралельних текстів «Українська словесна мова – Українська жестова мова». На основі розробленої математичної моделі граматично доповненої онтології української словесної та жестової мов для побудови системи перекладу УЖМ розроблено **предметно-орієнтовану мову (ПОМ)** опису для наповнення цієї граматично доповненої онтології.

Для вирішення проблеми лексичної багатозначності слів укладено словник концептів «Українська словесна мова – Українська жестова мова». Під концептом розуміємо значення жестового або словесного висловлювань, які виражають те саме поняття (процес, дію, ознаку), і можуть бути скорочені лише, коли слово або жест відновлюється з контексту. Наприклад, словосполучення «годинник йде» вважаємо концептом, тому що в жестовій мові для показу цього словосполучення використовують один жест «ГОДИННИК\_ЙДЕ», а словосполучення «людина йде» – не вважатимемо концептом (використовують жести «ЛЮДИНА» та «ЙТИ»).

Для якісного перекладу української жестової мови укладено перелік концептів, ці концепти згруповано та розроблено на їхній основі онтологію



українських словесної та жестової мов, яка містить інформацію про граматичні зв'язки між поняттями. Розроблення такої граматично доповненої онтології та її використання в системі машинного перекладу української жестової мови дає змогу частково вирішити проблему усунення багатозначності слів та підвищити оцінку якості перекладу.

Математична модель онтології визначена як кортеж  $O = \langle C, M, R_c \rangle$ , де  $C = \{c_i\}$  – набір понять (концептів) предметної галузі,  $M_i = \{m_{i1}, \dots, m_{di}\}$  – множина атрибутів поняття  $c_i$  ( $d$  – кількість атрибутів, що описують це поняття),  $R_c$  – відношення між концептами (гіпонімії, гіперонімії, меронімії, голонімії та ін.).

Означимо математичну модель граматично доповненої онтології як кортеж

$$O_G = \langle O, P, E, T, R_p \rangle, \quad (1)$$

де  $O$  – онтологія,  $P = \{p_i\}_{i=1,n}$  – набір предикатів;

$E = \{e_i\}$  – набір висловлювань, кожне висловлювання  $e_i = ((w_1, g_1), (w_2, g_2), \dots, (w_n, g_n))$  – кортеж з граматично доповнених термінів;

$T = \{t_j\}$  – набір параметризованих висловлювань,  $t_j = (e_j, f_j, p_j)$  – трійка, що містить висловлювання  $e_j$ , функцію параметризації  $f_j: \{1, 2, \dots, \text{Len}(e_j)\} \rightarrow \{0, 1, \dots, N(p_j)\}$  та предикат  $p_j$ ,  $\text{Len}(e_j)$  позначає довжину кортежу  $e_j$ ,  $N(p_j)$  – кількість місць предиката  $p_j$ ;

$R_p$  – відношення, яке ставить у відповідність предикати і дієслівні концепти (синсети).

Для деякого предиката  $p_j$  і деякого висловлювання  $e_j$  функція параметризації  $f_j(k)$  визначена як 0 для терміна в позиції  $k$  висловлювання  $e_j$ . Значення  $f(k) > 0$  означає, що відповідний термін у положенні  $k$  є аргументом предиката з порядковим номером  $f(k)$ , і може бути замінений іншим терміном з множини гіпонімів терміна  $w_k$ . Якщо відповідний предикат містить  $n$  місць і для кожного  $i \in \{1, 2, \dots, n\}$  існує  $k \in \{1, 2, \dots, \text{Len}(e_j)\}$  таке, що  $f(k) = i$ , то висловлювання  $e_j$  повністю визначає предикат  $p_j$ . В іншому випадку, деякі аргументи предиката вважаються невизначеними у реченні. Вони можуть бути зовсім невідомі або можуть бути відновлені з контексту речення.

Для практичного втілення розробленої математичної моделі граматично доповненої онтології необхідно створити методи та засоби, які б давали змогу описувати цю онтологію, і також перетворювали її на форму, придатну для використання у системах машинного перекладу.

На основі розробленої математичної моделі граматично доповненої онтології української словесної та жестової мов для побудови системи перекладу УЖМ розроблено предметно-орієнтовану мову опису для наповнення цієї граматично доповненої онтології.

Предметно-орієнтована мова UkrNet реалізована з використанням додатка Xtext середовища розроблення програмного забезпечення Eclipse, який дає змогу

для граматики створити ПОМ. Платформа Xtext на основі граматики предметно-орієнтованої мови забезпечує створення набору засобів для наповнення, редагування та верифікації граматично доповненої онтології, що містить аналізатор коду мови, інструмент форматування коду, компілятор, редактор коду та ін.

Розроблення предметно-орієнтованої мови UkrNet здійснено з урахуванням граматики цієї мови в Xtext. Потім, використовуючи засоби Xtext, створено синтаксичний аналізатор і текстовий редактор для мови UkrNet.

Для того, щоб забезпечити можливість введення в онтологію граматичних категорій розроблено спеціалізовану мову опису граматично доповненої онтології. Розглянемо опис понять, за допомогою яких описано цю онтологію.

Для позначення граматичних значень введено поняття *grammatical\_attribute* <назва атрибуту> [= onuc]. Для опису граматично доповненої онтології введено такі граматичні значення:

**grammatical\_attribute** ім\_наз // іменник у називному відмінку  
**grammatical\_attribute** ім\_род // іменник у родовому відмінку  
**grammatical\_attribute** ім\_дав // іменник у давальному відмінку  
**grammatical\_attribute** ім\_знах // іменник у знахідному відмінку  
**grammatical\_attribute** ім\_оруд // іменник у орудному відмінку  
**grammatical\_attribute** ім\_місц // іменник у місцевому відмінку  
**grammatical\_attribute** д\_інф // дієслово (інфінітив)  
**grammatical\_attribute** д\_жест // дієслово (жест)  
**grammatical\_attribute** ім\_жест // іменник (жест)

Для позначення іменникових синсетів використовуємо такий опис *synset* <назва синсету> [= onuc]. Кожен синсет відповідає деякому значенню, поняттю мови. Синсет можна розглядати як множину значень слів, які виражають одне поняття. Усі іменники, дієслова, прикметники і прислівники української словесної та української жестової мов утворюють синонімічні набори (синсети), кожен з яких подає один семантичний концепт. Наприклад:

**synset** сутність  
**synset** людина\_особа

Для позначення дієслівних синсетів використовують такий опис *synset* <назва синсету> [= onuc] (<newline> висловлювання)\* (<newline> onuc доменів)\*. Наприклад:

**synset** учити\_навчати = передавати знання  
 учити 1.ім\_знах 2.ім\_дав  
 давати поради 1.ім\_дав з 2.ім\_род  
 жест навчати 1.ім\_жест 2.ім\_жест  
 0: може\_вчити  
 1: може\_вчитись  
 2: містить\_знання

Цифри використовують для встановлення зв'язків між граматично доповненими термінами та предметними змінними предикату. У наведеному вище прикладі синсету «вчити» відповідає предикат *вчити(a,b,c)*, де *a* – це хтось, хто вчить, *b* – це хтось, хто вчиться, і *c* – певні знання.

Основним відношенням, яке використовують для синтаксично-семантичного розбору речень, є відношення гіперонімії та обернене йому відношення гіпонімії. Для позначення цього відношення використовують такий опис *hypernym* ( $\langle \text{синсет1} \rangle = \langle \text{синсет2} \rangle$ ). Наприклад:

*hypernym* (книга) = містить\_знання

Крім того, введено відношення асоціації та меронімії/голонімії (частина-ціле). Ці відношення мають такий опис:

*parts*( $\langle \text{синсет1} \rangle$ )= $\langle \text{синсет2} \rangle$  [, $\langle \text{синсетN} \rangle$ ]\*

*optional\_parts*( $\langle \text{синсет1} \rangle$ )= $\langle \text{синсет2} \rangle$  [, $\langle \text{синсетN} \rangle$ ]\*

*associations*( $\langle \text{синсет1} \rangle$ )= $\langle \text{синсет2} \rangle$  [, $\langle \text{синсетN} \rangle$ ]\*

*optional\_associations*( $\langle \text{синсет1} \rangle$ )= $\langle \text{синсет2} \rangle$  [, $\langle \text{синсетN} \rangle$ ]\*

Ці відношення використовуються для підвищення ймовірності значення слова, яке пов'язане відношеннями асоціації або меронімії зі сусідніми словами. Наприклад:

*parts* (книга) = сторінка

*optional\_parts* (сторінка) = номер\_сторінки

*associations* (студент) = університет

*optional\_associations* (учитель) = школа

Для побудови граматично доповненої онтології вибрано такі предметні галузі «Навчання», «Природа», «Подорож», «Держава», «Сім'я», «Виробництво», «Професії», «Військо», «Театр», «Культура», «Лікарня». Для цього зібрано 1200 слів, які пов'язані з цими галузями та описано пояснення до кожного слова з використанням тлумачного словника української мови та словника жестів. До кожного слова також зібрано всі його синоніми. Після цього до редактора онтологій внесено ці слова у вигляді синсетів (синонімічних рядів) та описано граматичні конструкції, в яких використовується це слово. У синсет також введено граматичну конструкцію жестової мови. Так, ми зможемо отримати однозначний переклад певного словосполучення.

Розроблений метод наповнення онтології з використанням предметно-орієнтованих мов дав змогу розробити граматично доповнену онтологію української словесної та української жестової мов.

Застосування розробленої математичної моделі у парсері української словесної та жестової мов на основі ймовірнісних афіксних контекстно-вільних граматик (Affix probabilistic context free grammar, APCFG) дало змогу отримати синтаксично-семантичний розбір речень корпусу паралельних текстів «Українська словесна мова – Українська жестова мова» за умови відповідності речення правилам онтологічного словника. Загалом підхід, що ґрунтується на граматично доповненій онтології, дає змогу покращити результат перекладу української жестової та словесної мов і його можна ефективно використовувати у системах машинного перекладу.

У **третьому розділі** описано розроблені методи та засоби перекладу анотованої української жестової мови. Проведено застосовність таких методів машинного перекладу української жестової мови, як: статистичного методу, методу на основі правил та розробленого методу з використанням ГДО.

Застосовність статистичного методу машинного перекладу української словесної мови на українську жестову мову, а саме застосування моделі IBM № 1 та EM-алгоритму для вирівнювання слів, проведено на реченнях корпусу паралельних текстів «УСМ – УЖМ».

Систематизовано правила перекладу з української словесної мови на анотовану українську жестову та навпаки. Проведено тестування методу перекладу на основі правил для речень з корпусу паралельних текстів «УСМ – УЖМ».

На рис. 1 подано блок-схему алгоритму перекладу із використанням граматично доповненої онтології.

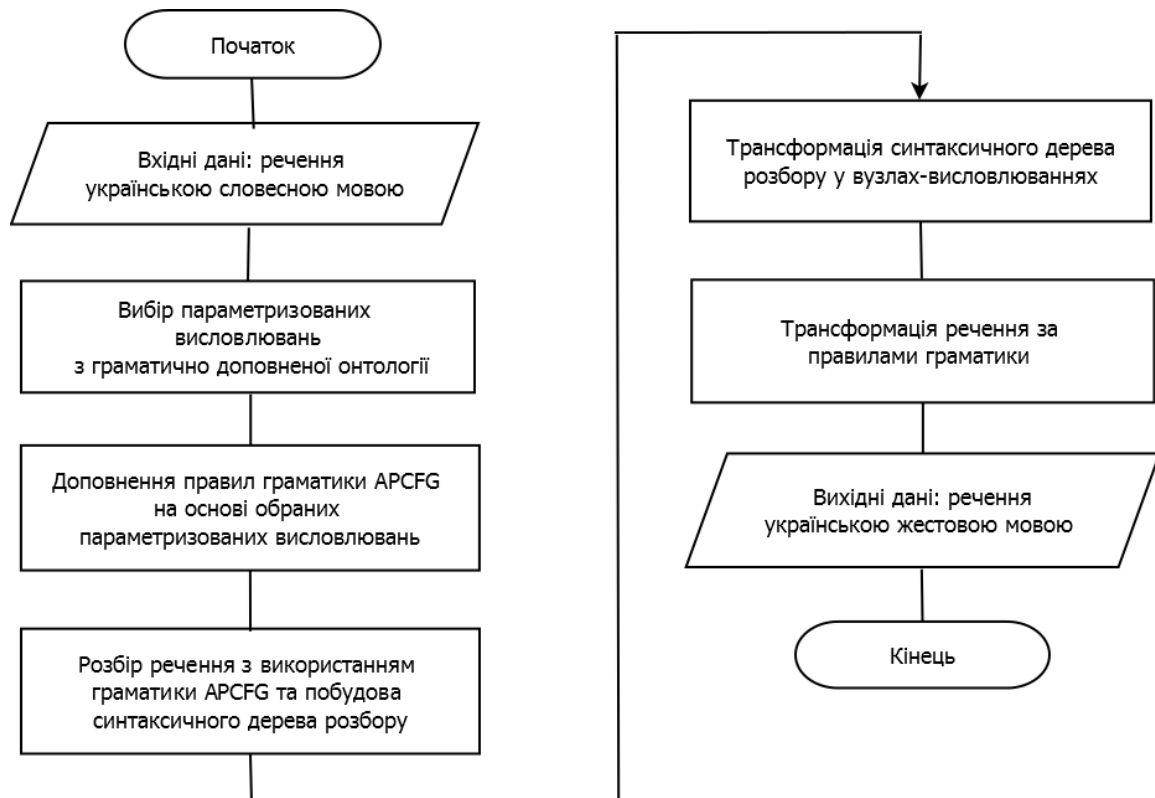


Рис. 1. Блок-схема алгоритму перекладу із застосуванням граматично доповненої онтології

На вхід подається речення українською словесною мовою. Далі програма здійснює пошук у граматично доповненій онтології та додає їх як правила виведення. Наступний етап полягає у розборі речення з використанням афіксної ймовірнісної контекстно-вільної граматики. У результаті розбору будується дерево семантично-синтаксичного розбору (дерево складових) речення (рис. 2, а), коли розбір відповідає граматично доповненій онтології (вага правила більша за 1).

За допомогою відношень між поняттями у граматично доповненій онтології на місця параметризованих змінних можна поставити лише правильні по змісту поняття. Відповідно до наведеної граматично доповненої онтології для речення «Я розповідаю столу казку» дерево синтаксичного розбору відповідає звичайному синтаксису речення (вага правила дорівнює 1) та зображено на рис. 2, б.

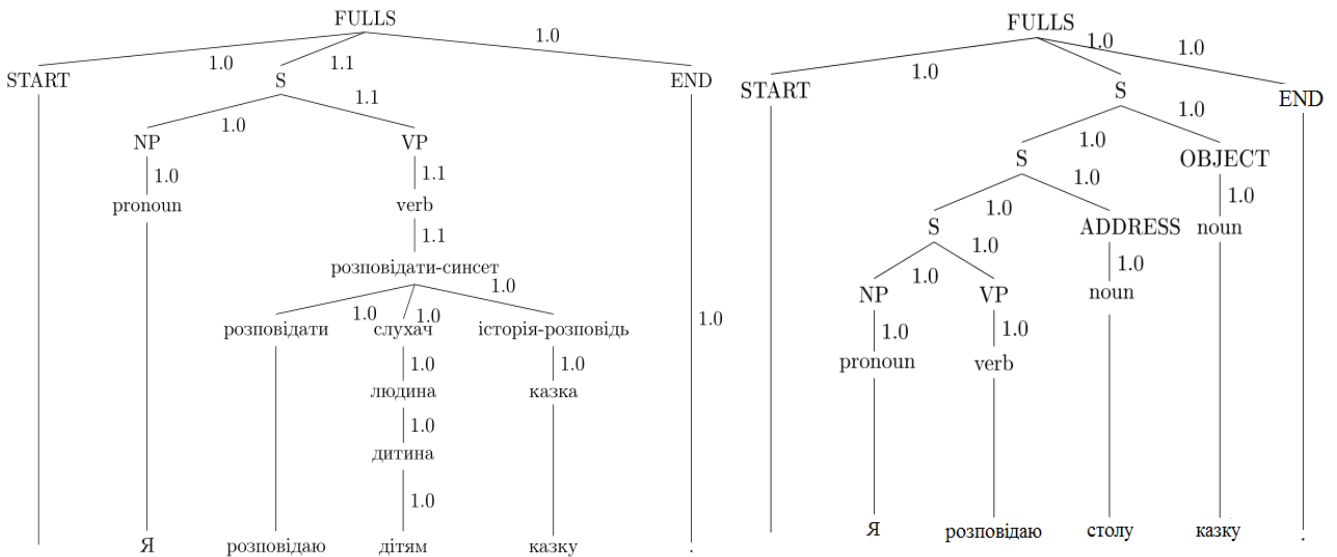


Рис. 2. Древа семантично-синтаксичного розбору для речень: а – «Я розповідаю дітям казку»; б – «Я розповідаю столу казку»

Усі правила, за якими здійснюється семантично-синтаксичний розбір речень УСМ, поділено на п'ять груп правил для:

- 1) врахування порядку слів у реченні;
- 2) іменникових синсетів;
- 3) дієслівних синсетів;
- 4) питальних речень;
- 5) речень, у яких порушено порядок слів.

Крім того, розроблено правила семантично-синтаксичного розбору речень анотованої української жестової мови. Ці правила поділяються на чотири групи (немає п'ятої групи правил, у яких порушено порядок слів, оскільки для УЖМ порядок слів у реченні чітко визначений) та аналогічні до перших чотирьох груп української словесної мови.

Правила семантики, які створені з онтології для дерева семантично-синтаксичного розбору (позначення наведено у табл. 1), зображеного на рис. 2, а:

$\langle \text{людина} \rangle [NCG] \rightarrow \langle \text{дитина} \rangle [NCG]$

$\langle \text{слухач} \rangle [NCG] \rightarrow \langle \text{людина} \rangle [NCG]$

$\langle \text{історія-розповідь} \rangle [NCG] \rightarrow \langle \text{казка} \rangle [NCG]$

$\langle \text{розповідати-синсет} \rangle [PN] * 1.1 \rightarrow \langle \text{розповідати} \rangle [PN] \langle \text{історія-розповідь} \rangle [c4]$   
 $\langle \text{слухач} \rangle [c3]?$

$\text{verb}[PN] \rightarrow \langle \text{розповідати-синсет} \rangle [PN]$

Таблиця 1

Позначення для дерев семантично-синтаксичного розбору

Позначення	Пояснення
<i>N</i>	число (number)
<i>C</i>	відмінок (casus)
<i>G</i>	рід (gender)
<i>P</i>	особа (person)

Можливість побудови дерев семантично-синтаксичного розбору для речень української словесної та анотованої жестової мов дає змогу зрозуміло та однозначно перекладати між цими мовами. Наприклад, для речення «Я розповідати діти казка» дерево розбору зображено на рис. 3.

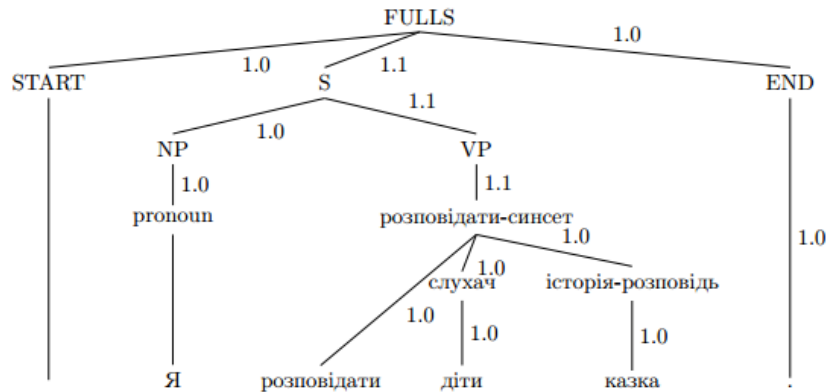


Рис. 3. Дерево семантично-синтаксичного розбору для речення УЖМ «Я розповідати діти казка»

Для анотованої української жестової мови правила семантики, які створені з онтології для речення УЖМ «Я розповідати діти казка»:

**<слухач>[NCG] → <діти>[NCG]**

**<історія-розповідь>[NCG] → <казка>[NCG]**

**<розповідати-синсет>[PN] \* 1.1 → <розповідати>[PN] <історія-розповідь> <слухач>?**

**verb[PN] → <розповідати-синсет>[PN]**

Після семантично-синтаксичного розбору речення дерево складових цього речення трансформується у дерево залежностей згідно з алгоритмом трансформації. Як для дерев граматики залежностей, так і для дерев граматики складових важливим є поняття «головної складової». Для дерев граматики складових головною складовою позначається головне слово у виразі і від цього слова залежать усі інші слова виразу. Алгоритм визначення головної складової в деревах граматики складових містить такі кроки:

- 1) головна складова вузла VP є кореневим вузлом (корінь) у дереві граматики залежностей;
- 2) усі вузли NP та NG, у яких немає вузла VP, є залежними від кореня;
- 3) якщо вузол S є батьком VP, то всі NP та NG, які є нащадками вузла S, також є залежними від цього кореня;
- 4) головними складовими вузлів NP, NG, OBJECT, ADDRESS, додатка DNP, є іменник (noun) або займенник (pronoun); усі решта вузлів, які входять до NP, NG, OBJECT, ADDRESS, DNP, зокрема прикметники (adj) є залежними від них;
- 5) головною складовою вузла VP є дієслово (verb); якщо вузол VP містить допоміжне дієслово (IVP), то воно є залежним від дієслова verb;
- 6) головною складовою вузлів TARGET, FROM є додаток (DNP);
- 7) якщо вузол-нащадок S є нащадком вузла-батька S, тоді голова вузла-нащадка S буде головою вузла-батька S.

Алгоритм трансформації визначає відношення нащадок-батько для вузлів дерева граматики складових і перетворює його на дерево граматики залежностей.

Просуваючись по відношенням нащадок-батько від вузлів-листіків до вузла-кореня дерева, ми можемо позначити кожен вузол тегом «головна складова». Так, можна побудувати дерева граматики залежностей.

Для перекладу речень УСМ у речення анотованої УЖМ та навпаки розроблено правила трансформації дерев граматики залежностей за заданими правилами граматики УЖМ. На цьому етапі встановлюємо визначений відповідно до УЖМ порядок слів у реченні. Для перетворення речення УСМ на речення українською жестовою мовою розроблено спеціальну систему правил. Система складається з правил перекладу на жестову мову, правил порядку слів у реченнях відповідно до граматики УЖМ.

Розроблені методи для семантично-синтаксичного розбору речень, методи використання граматично доповненої онтології для перекладу анотованої української жестової мови та правила перекладу в сукупності є основними елементами системи перекладу УЖМ, описаної у дисертаційній роботі.

У **четвертому розділі** спроектовано архітектуру програмно-алгоритмічного комплексу перекладу анотованої української жестової мови “SignTranslator”, схему функціонування якого зображено на рис. 4. Наведено результати тестування розробленого програмно-алгоритмічного комплексу.

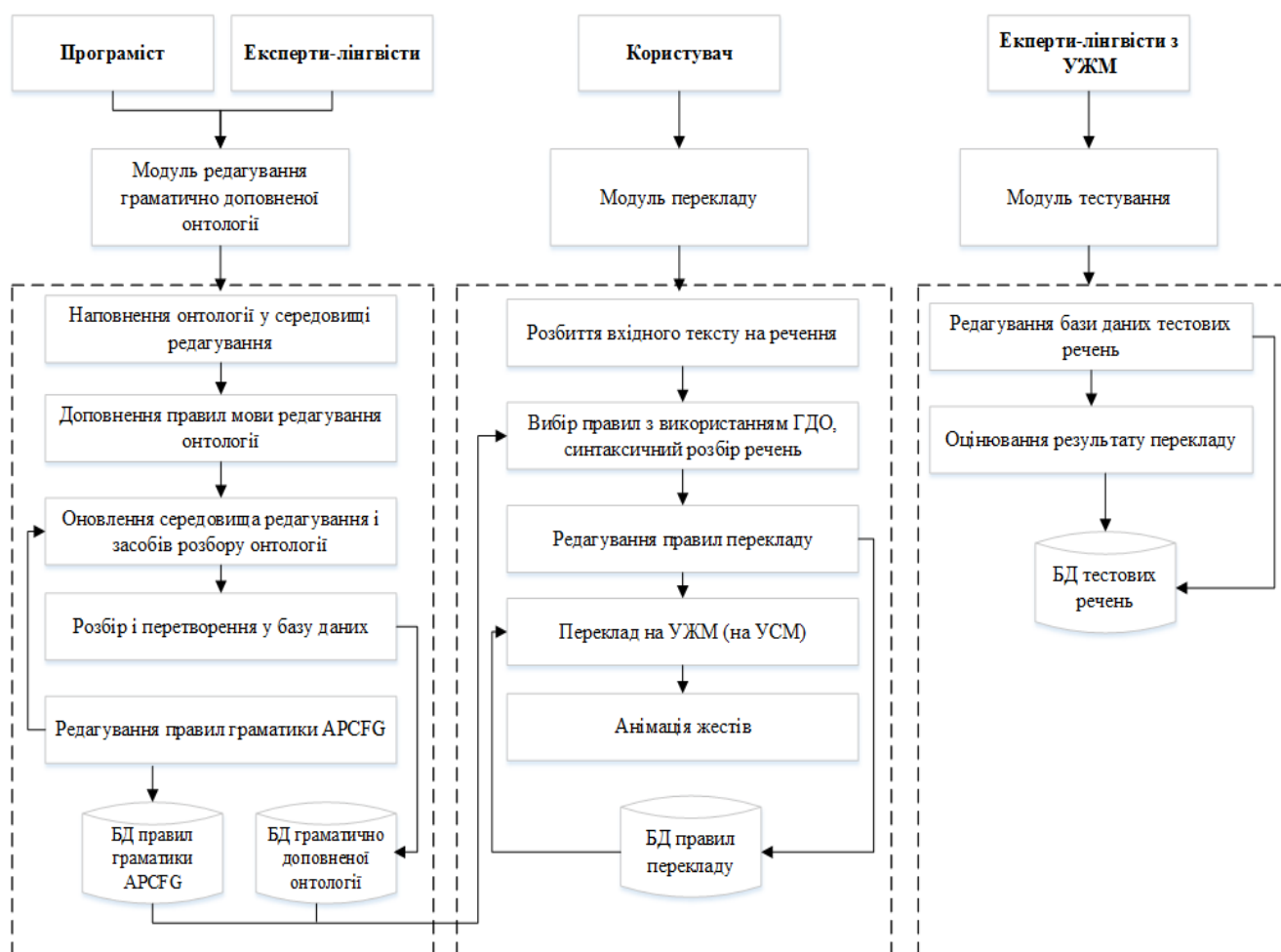


Рис. 4. Схема функціонування програмно-алгоритмічного комплексу перекладу анотованої української жестової мови “SignTranslator”

В основу архітектурних рішень під час побудови програмно-алгоритмічного комплексу перекладу анотованої української жестової мови покладено модульну структуру, що дало змогу реалізувати його у вигляді окремих системно пов'язаних функціональних модулів. Кожен модуль працює з певним вхідним набором даних. Розроблено такі функціональні програмно-алгоритмічні модулі:

- модуль наповнення граматично доповненої онтології;
- модуль перекладу із застосуванням правил та онтології;
- модуль тестування перекладу.

Модуль наповнення граматично доповненої онтології реалізує такі процеси: оновлення середовища редагування та засобів розбору онтології, розбір онтології та її перетворення на базу даних, доповнення правил граматики APCFG та правил перекладу.

Модуль тестування перекладу реалізує процеси оцінювання результату перекладу та доповнення бази тестових речень.

Модуль перекладу із застосуванням правил та онтології реалізує такі процеси: розбиття вхідного тексту на речення та визначення їхньої тематики, семантично-синтаксичний розбір речень та переклад на анотовану українську жестову мову.

Програмно-алгоритмічний комплекс реалізовано за допомогою програмного середовища Eclipse (модуль перекладу та модуль тестування). Модуль наповнення граматично доповненої онтології розроблено з використанням додатка XText середовища Eclipse.

Схема роботи програмно-алгоритмічного комплексу перекладу української жестової мови, зображена на рис. 4, передбачає наявність чотирьох баз даних: БД граматично доповненої онтології, БД правил граматики APCFG, БД правил перекладу, БД тестових речень.

Інфологічну модель «сутність–зв'язок» граматично доповненої онтології української словесної та жестової мов зображено на рис. 5.

Для оцінювання якості системи перекладу анотованої української жестової мови проведено чотири експерименти, які полягали у перекладі тестових текстів. Усі тестування проводились на 220 реченнях корпусу паралельних текстів «Українська словесна мова – Українська жестова мова».

У першому експерименті для перекладу застосовувались лише статистичні методи перекладу.

У другому експерименті для перекладу використовувались 20 правил перекладу з української словесної мови на українську жестову та навпаки.

Третій експеримент полягав у тестуванні системи комп'ютерного перекладу УЖМ на основі правил із використанням словника концептів. Словник концептів містить 50 жестових концептів та їхніх відповідних перекладів українською словесною мовою. Для тестування взято 20 правил перекладу та словник концептів, які застосовано для перекладу 220 речень.

Четвертий експеримент проводився із застосуванням 281 правил граматики APCFG (154 правил для УСМ та 127 правил для УЖМ) та бази граматично доповненої онтології.



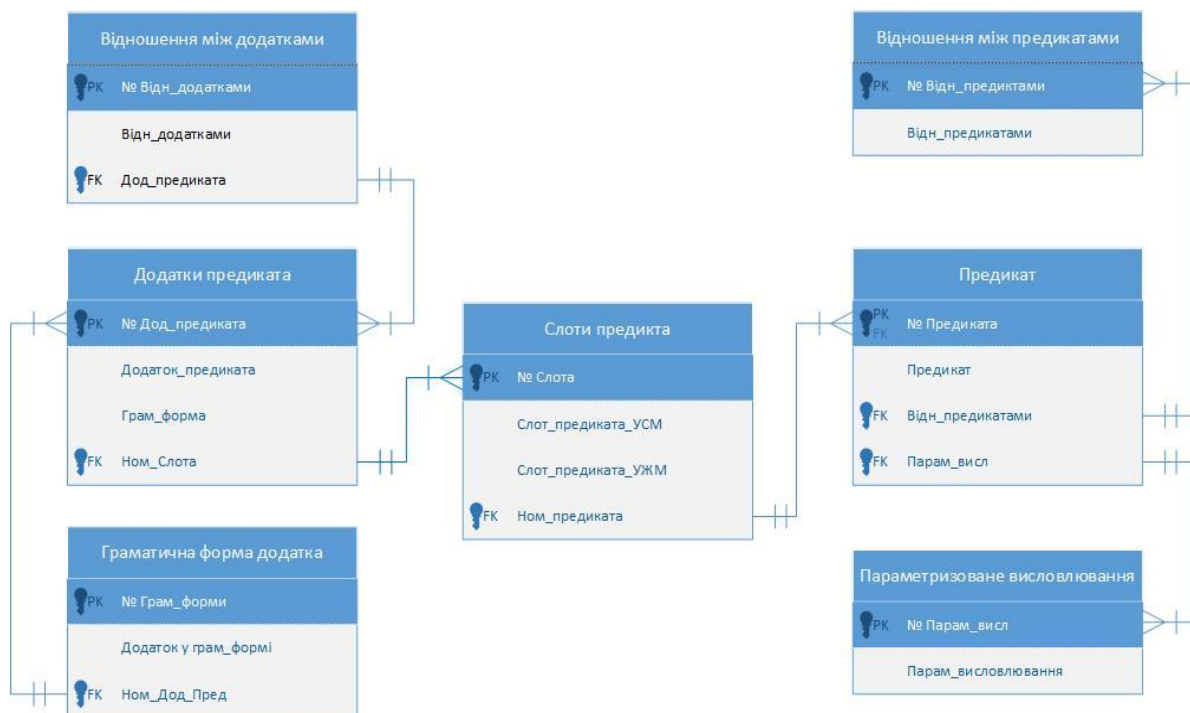


Рис. 5. Інфологічна модель «сутність–зв’язок» граматично доповненої онтології української словесної та жестової мов

Для аналізу ефективності розроблених методів запропоновано показник правильно перекладених речень:

$$C = \frac{N_{correct\_trans}}{N_{general\_trans}} \times 100\%, \quad (2)$$

де  $N_{correct\_trans}$  – кількість правильно перекладених тестових речень,  $N_{general\_trans}$  – загальна кількість тестових речень.

Результати оцінювання якості перекладу анотованої української жестової мови наведено у табл. 2 та на рис. 6.

Таблиця 2

Оцінка якості перекладу (%)

	Експеримент № 1	Експеримент № 2	Експеримент № 3	Експеримент № 4
<b>УСМ→УЖМ</b>	57,2	75,3	85,4	93,2
<b>УЖМ→УСМ</b>	50,5	68,7	81,6	91,4

Проведене оцінювання якості системи перекладу з використанням граматично доповненої онтології порівняно із статистичним методом, методом перекладу на основі правил та методом з використанням словника концептів «УСМ–УЖМ» дало змогу покращити відсоток правильно перекладених речень до 93,2 %.

Під час виконання дисертаційного дослідження один із основних акцентів зроблено на створенні граматично доповненої онтології української словесної та жестової мов для її використання в системах комп’ютерного перекладу.

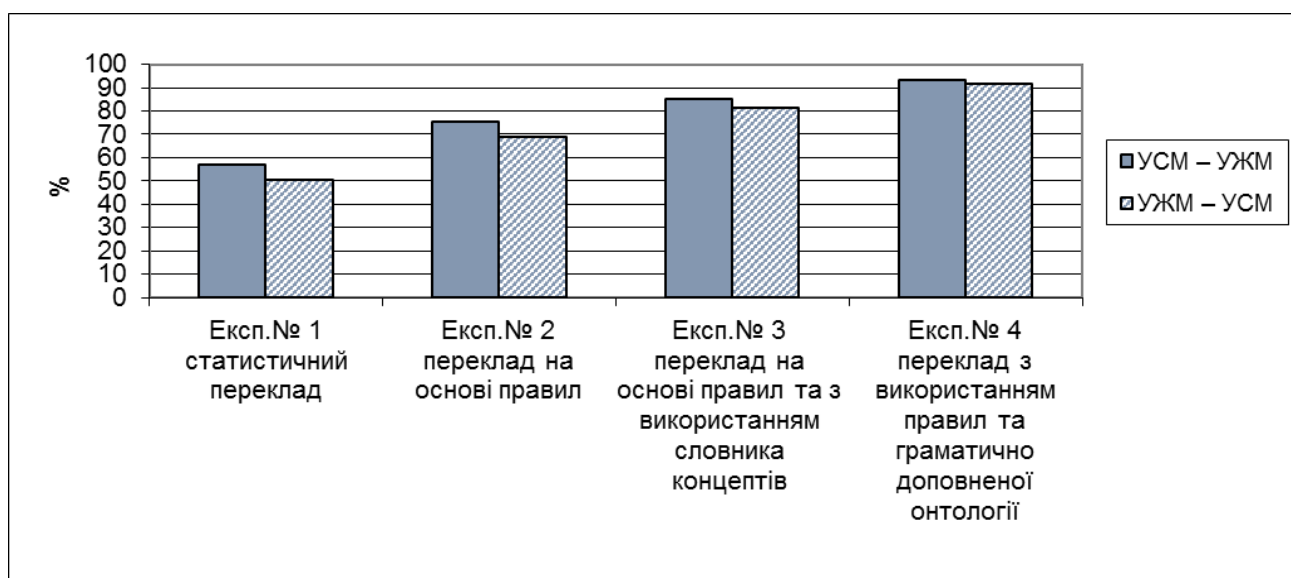


Рис. 6. Порівняння методів перекладу анотованої української жестової мови

## ОСНОВНІ РЕЗУЛЬТАТИ ТА ВИСНОВКИ

У дисертаційній роботі вирішено актуальне науково-прикладне завдання комп'ютерного перекладу анотованої української жестової мови.

Для цього розроблено програмно-алгоритмічний комплекс перекладу анотованої української жестової мови на основі граматично доповненої онтології.

Отримано такі наукові результати:

1) вперше запропоновано архітектуру програмно-алгоритмічного комплексу перекладу анотованої української жестової мови, яка відрізняється від відомих іноземних аналогів додаванням модуля перекладу на основі граматично доповненої онтології, що дало змогу автоматично перекладати УЖМ;

2) вперше напрацьовано словник концептів «Українська словесна мова – Українська жестова мова» для перекладу анотованої української жестової мови, що дало змогу зменшити відсоток неправильно перекладених тестових речень на 8,25 %;

3) вперше розроблено систему правил перекладу з української словесної мови на анотовану українську жестову мову з використанням граматично доповненої онтології, що дало змогу підвищити якість перекладу з 57,2 % до 93,2 %;

4) вперше сформульовано правила семантично-синтаксичного розбору речень для української словесної та анотованої української жестової мов, що дало змогу здійснити семантичний аналіз речень для обох мов;

5) одержав подальший розвиток метод наповнення онтологій з використанням предметно-орієнтованих мов, що дало змогу розробити граматично доповнену онтологію;

6) удосконалено метод комп'ютерного перекладу на основі правил анотованої української жестової мови з використанням граматично доповненої онтології.

## СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Davydov M. Spoken and sign language processing using grammatically augmented ontology / M. Davydov, O. Lozynska // *Applied Computer Science. ACS journal.* – Poland, 2015. – Vol. 11. – No 2. – P. 29–42.
2. Lozynska O. Information technology for Ukrainian Sign Language translation based on ontologies / O. Lozynska, M. Davydov // *Econtechmod. An international quarterly journal.* – Lublin, 2015. – Vol. 4. – No 2. – P. 13–18.
3. Лозинська О. В. Математична модель граматично-доповненої онтології / О. В. Лозинська, М. В. Давидов // *Вісник Нац. техн. ун-ту «Харків. політех. ін-т».* – Харків : Нац. техн. ун-т «Харків. політех. ін-т», 2015. – № 11(1120). – С. 102–108.
4. Лозинська О. В. Предметно-орієнтована мова опису граматично-доповненої онтології / О. В. Лозинська, М. В. Давидов // *Управляющие системы и машины.* – Київ, 2015. – № 4. – С. 31–40.
5. Давидов М. В. Інфологічна модель концептів української жестової мови / М. В. Давидов, О. В. Лозинська, В. В. Пасічник // *Вісник Нац. ун-ту «Львівська політехніка» : «Інформаційні системи та мережі».* – Львів, 2014. – № 805. – С. 279–289.
6. Лозинська О. В. Машинний переклад на основі правил для перекладу на українську жестову мову / О. В. Лозинська, М. В. Давидов, В. В. Пасічник // *Інформаційні технології та комп'ютерна інженерія : Міжнар. наук.-техн. журн.* – Вінниця, 2014. – Т. 1. – № 29. – С. 11–17.
7. Лозинська О. В. Застосування методу машинного перекладу на основі правил для перекладу з української жестової мови / О. В. Лозинська // *Вісник Нац. ун-ту «Львівська політехніка» : Інформаційні системи та мережі.* – Львів, 2014. – № 783. – С. 368–375.
8. Лозинська О. В. Розроблення системи статистичного машинного перекладу української жестової мови на українську словесну мову для підтримки бізнес-процесів / О. В. Лозинська, А. С. Мельник, М. В. Давидов, Ю. В. Нікольський // *Радіоелектроніка та інформатика : наук.-техн. журн.* – Харків : Харків. нац. ун-т радіоелектроніки, 2013. – № 1 (60). – С. 35–39.
9. Лозинська О. В. Побудова системи правил для комп'ютерного перекладу української жестової мови на основі аналізу її синтаксичних конструкцій / О. В. Лозинська, М. В. Давидов // *Вісник Нац. ун-ту «Львівська політехніка» : Інформаційні системи та мережі.* – Львів, 2015. – № 814. – С. 319–326.
10. Лозинская О. В. Информационные технологии для незрячих и неслышающих людей / О. В. Лозинська, А. Б. Демчук // *Science and World : Междунар. науч. журнал.* – Волгоград, 2014. – № 1(5). – С. 102–105.
11. Лозинська О. В. Використання граматично-доповненої онтології для перекладу української жестової мови / О. В. Лозинська // *Математика. Інформаційні технології. Освіта : Збірник статей.* – Луцьк : Східноєвропейський національний ун-т ім. Лесі Українки, 2015. – № 2. – С. 108–114.
12. Давидов М. В. Освітні інформаційні технології для глухих та слабкочуючих. Напрацювання та досвід Національного університету «Львівська політехніка» : препринт / М. В. Давидов, О. В. Лозинська, О. В. Вербич. – Львів : Видавництво Львівської політехніки, 2013. – 48 с.

13. Lozynska O. Using rule-based machine translation for Ukrainian Sign Language / O. Lozynska, M. Davydov // Perspective Technologies and Methods in MEMS Design 'MEMSTECH-2014': proc. of the X Intern. Conf., 22–24 June 2014, Lviv, Ukraine / Min. of Education, Science, Youth and Sports of Ukraine, Lviv Politechnik National University. – Lviv, 2014. – P. 123–124.
14. Лозинська О. В. Використання дерев синтаксичного розбору для перекладу української жестової мови / О. В. Лозинська, М. В. Давидов, А. Б. Демчук // Матеріали Міжнар. наук.-практ. конф. «Інформаційні управляючі системи та технології». – Одеса : Одес. нац. морський ун-т, 2013. – С. 321–324.
15. Лозинська О. В. Інформаційна технологія перекладу української жестової мови з використанням концептів / О. В. Лозинська, М. В. Давидов // Матер. V Міжнар. наук.-техн. конф. «Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління». – Полтава : Полтав. нац. техн. ун-т ім. Ю. Кондратюка, 2014. – С. 45–49.
16. Lozynska O. Creating Ukrainian Sign Language corpus for its computer translation / O. Lozynska, M. Davydov, I. Nikolski // Perspective Technologies and Methods in MEMS Design 'MEMSTECH-2013' : proc. of the IX Intern. Conf., 16–20 April 2013, Polyana, Ukraine / Min. of Education, Science, Youth and Sports of Ukraine, Lviv Politechnik National University. – Lviv, 2013. – P. 155.
17. Grimnak O. Investigation of Ukrainian Sign Language to building effective methods of its computer translation / O. Grimnak, M. Davydov // Computer Science and Information Technologies 'CSIT-2012' : proc. of the VII Intern. Sci. and Techn. Conf., 20–24 Nov. 2012, Lviv, Ukraine / Min. of Education, Science, Youth and Sports of Ukraine, Lviv Politechnik National University. – Lviv, 2012. – P. 23.
18. Лозинська О. В. Методи машинного перекладу української жестової мови / О. В. Лозинська // Матер. IV Міжнар. наук.-практ. конф. «Математика. Інформаційні технології. Освіта». – Луцьк :, Східноєвропейський національний ун-т ім. Лесі Українки, 2015. – С. 38–40.
19. Демчук А. Б. Тифлокоментування: зародження та перспективи розвитку / А. Б. Демчук, О. В. Лозинська // Матер. Міжнар. конф. «Наука как основа мирного диалога». – Донецьк, 2014. – С. 17–20.
20. Демчук А. Б. Основні прийоми і способи створення тифлокоментарів / А. Б. Демчук, О. В. Лозинська, М. Я. Голяк // Матер. Міжнар. наук.-практ. конф. молодих вчених та студентів «Інформаційні технології, економіка та право: стан та перспективи розвитку». – Чернівці : ПВНЗ «Буковинський університет», 2014. – С. 4–5.

## АНОТАЦІЇ

**Лозинська О. В. Система комп'ютерного перекладу української жестової мови з використанням граматично доповненої онтології.** – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 10.02.21 – «Структурна, прикладна та математична лінгвістика». – Національний університет «Львівська політехніка» Міністерства освіти і науки України, Львів, 2016.

У дисертації розроблено методи та засоби системи комп'ютерного перекладу анотованої української жестової мови з використанням граматично доповненої онтології. Досліджено проблему перекладу анотованої української жестової мови. Вперше побудовано систему правил для перекладу анотованої української жестової мови із використанням граматично доповненої онтології. Удосконалено метод комп'ютерного перекладу на основі правил анотованої української жестової мови з використанням граматично доповненої онтології. Одержав подальший розвиток метод наповнення онтологій з використанням предметно-орієнтованих мов, що дало змогу розробити граматично доповнену онтологію. У результаті розроблено систему комп'ютерного перекладу анотованої української жестової мови, що дало змогу підвищити відсоток правильно перекладених речень української словесної мови на українську жестову до 93,2 % порівняно зі статистичним методом перекладу (57,2 %) та методом перекладу на основі правил (75,3 %).

*Ключові слова:* система комп'ютерного перекладу, українська жести́мова мова, граматично доповнена онтологія, дерево синтаксичного розбору, афіксна ймовірнісна контекстно-вільна грамати́ка.

**Лозинская О.В. Система компьютерного перевода украинского жестового языка с использованием грамматически дополненной онтологии.** – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 10.02.21 – «Структурная, прикладная и математическая лингвистика». – Национальный университет «Львівська політехніка» Министерства образования и науки Украины, Львов, 2016.

В диссертации разработаны методы и средства системы компьютерного перевода аннотированного украинского жестового языка с использованием грамматически дополненной онтологии. Исследована проблема перевода аннотированного украинского жестового языка. Впервые построена система правил для перевода аннотированного украинского жестового языка с использованием грамматически дополненной онтологии. Усовершенствован метод компьютерного перевода на основе правил аннотированного украинского жестового языка с использованием грамматически дополненной онтологии. Получил дальнейшее развитие метод наполнения онтологий с использованием предметно-ориентированных языков, что позволило разработать грамматически дополненную онтологию. В результате разработана система компьютерного перевода аннотированного украинского жестового языка, что позволило повысить процент правильно переведенных предложений украинского словесного языка на украинский жестовый язык до 93,2 % по сравнению со статистическим методом перевода (57,2 %) и методом перевода на основе правил (75,3 %).

*Ключевые слова:* система компьютерного перевода, украинский жестовый язык, грамматически дополненная онтология, дерево синтаксического разбора, аффиксная вероятностная контекстно-свободная грамматика.

**Lozynska O. V. Computer translation system of Ukrainian sign language using grammatically augmented ontology. – Manuscript.**

Thesis for a candidate's degree in technical sciences by specialty 02.10.21 – Structural, applied and mathematical linguistics. – Lviv Polytechnic National University Ministry of Education and Science of Ukraine, Lviv, 2016.

In the thesis an important scientific task of developing computer and linguistic methods and means for computer translation of annotated Ukrainian sign language using grammatically augmented ontology was resolved.

The problem of developing a machine translation system for sign language has been studied by scientists for a long time. The solution to this problem can provide new communication opportunities for people with hearing impairments. The challenge of translation from Ukrainian sign language (USL) to Ukrainian spoken language (USpL) refers to tasks of machine translation.

Author performed the original research of the linguistic peculiarities of Ukrainian sign language and developed the grammatically augmented ontology (GAO) of Ukrainian spoken language and Ukrainian sign language. The mathematical model of grammatically augmented ontology was developed. The model provides the possibility of integrating expressions into ontologies and supports the means for description of grammatical attributes.

Author developed software and algorithmic complex for bidirectional translation Ukrainian sign language. Method and algorithm for machine translation annotated Ukrainian sign language based on grammatically augmented ontology were developed.

The use of the developed grammatically augmented ontology for parsing sentences in Ukrainian spoken and Ukrainian sign languages improved the performance of APCFG parser. The major increase in percentage of correctly parsed sentences was achieved for Ukrainian sign language. The rules of syntax and semantic analysis of USL and USpL sentences based on grammatically augmented ontology were developed.

Domain specific language named GAODL for description of grammatically augmented ontology was created. The grammar of the language was defined by means of Xtext framework for Eclipse. This approach lets to use Eclipse environment with intelligent code completion for editing ontology files. The use of GAODL facilitates uniform editing and processing of grammatically augmented ontologies. These ontologies could be created for specific subject areas and lately merged to obtain upper ontologies. The GAODL language contains constructions for definition of new grammatical attributes, synsets, relations on synsets, predicates and expressions.

As a result, a computer translation system of annotated Ukrainian sign language using GAO was developed. The quality of Ukrainian sign language translation was increased to 93.2 % compared to the statistical method of translation (57.2 %) and the rule-based method of translation (75.3 %).

The main results of the dissertation research implemented in Lviv Maria Pokrova Secondary Residential School for Deaf Children.

*Key words:* computer translation system, Ukrainian sign language, grammatically augmented ontology, parsing trees, affix probabilistic context free grammar.