

Міністерство освіти і науки України
Національний університет «Львівська політехніка»

На правах рукопису

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

Дисертація на здобуття наукового ступеня
кандидата технічних наук

Науковий керівник: доктор технічних наук, професор,
лауреат державної премії України в
галузі науки та техніки,
Пасічник Володимир Володимирович
Національний університет «Львівська
політехніка», професор кафедри
інформаційних систем та мереж

Львів – 2016

ЗМІСТ

ВСТУП	5
РОЗДІЛ 1. ОГЛЯД ВІДОМИХ ДОСЛІДЖЕНЬ ЖЕСТОВОЇ МОВИ	13
1.1. Відомі методи та засоби перекладу жестових мов	13
1.2. Методи запису жестової мови.....	24
1.3. Подання словників жестової мови в комп'ютеризованих системах.....	27
1.3.1. Корпуси паралельних текстів	27
1.3.2. Онтологічні словники	29
1.4. Правила перекладу жестової мови	32
Висновки до першого розділу	35
РОЗДІЛ 2. ГРАМАТИЧНО ДОПОВНЕНА ОНТОЛОГІЯ ДЛЯ УКРАЇНСЬКОЇ СЛОВЕСНОЇ ТА УКРАЇНСЬКОЇ ЖЕСТОВОЇ МОВ	37
2.1. Дослідження відношень на множині концептів «УСМ – УЖМ»	37
2.2. Інфологічна модель словника відношень між концептами та їхнім поданням в українській словесній та жестовій мовах.....	44
2.3. Математична модель граматично доповненої онтології.....	55
2.4. Предметно-орієнтована мова опису граматично доповненої онтології.....	59
2.4.1. Формальне подання мови опису граматично доповненої онтології....	61
Висновки до другого розділу	66
РОЗДІЛ 3. МЕТОДИ ДВОСТОРОННЬОГО ПЕРЕКЛАДУ АНОТОВАНОЇ УКРАЇНСЬКОЇ ЖЕСТОВОЇ МОВИ	68
3.1. Застосування статистичних методів машинного перекладу анотованої української жестової мови.....	68
3.2. Застосування методів машинного перекладу анотованої УЖМ на основі правил	74
3.2.1. Правила перекладу речень української словесної мови на українську жестову	74
3.2.2. Правила перекладу речень української жестової мови на українську словесну.....	82

3.3. Застосування методу машинного перекладу анотованої української жестової мови з використанням ГДО.....	88
3.3.1. Трансформація дерева складових у дерево залежностей.....	100
Висновки до третього розділу.....	107
РОЗДІЛ 4. ПРОГРАМНО-АЛГОРИТМІЧНИЙ КОМПЛЕКС ПЕРЕКЛАДУ АНОВОВАНОЇ УКРАЇНСЬКОЇ ЖЕСТОВОЇ МОВИ.....	109
4.1. Схема роботи програмно-алгоритмічного комплексу перекладу анотованої української жестової мови.....	109
4.1.1. Інструкція користувачеві.....	114
4.2. Структура баз даних програмно-алгоритмічного комплексу.....	118
4.2.1. База даних граматично доповненої онтології	118
4.2.2. База даних правил перекладу	120
4.2.3. База даних правил граматики APCFG.....	122
4.2.4. База даних тестових речень.....	124
4.3. Аналіз результатів тестування системи комп'ютерного перекладу української жестової мови.....	126
Висновки до четвертого розділу.....	130
ВИСНОВКИ	132
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	134
ДОДАТОК А	147
ДОДАТОК Б.....	152
ДОДАТОК В	156

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

СМП – статистичний машинний переклад

ПАК – програмно-алгоритмічний комплекс

ЖМ – жестова мова

УСМ – українська словесна мова

УЖМ – українська жестова мова

ГДО – граматично доповнена онтологія

ПОМ – предметно-орієнтована мова

APCFG – афіксна ймовірнісна контекстно-вільна граматики

БД – база даних

ВСТУП

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 поліграфічний 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.

Українська жестова мова є природним способом спілкування слабкочуючих та нечуючих людей та є невід'ємною частиною їхньої особистості. Для вирішення проблеми доступу до інформаційних ресурсів такої категорії громадян розробляють принципово нові комп'ютерні технології доступу до освітніх інформаційних ресурсів, облаштовують сучасні комп'ютеризовані робочі місця, розробляють спеціальне програмне забезпечення, яке допомагає у вивченні української жестової мови та перекладі УСМ на УЖМ і навпаки. Такі розроблення, крім суто технічного комп'ютерного застосування, мають велике соціальне значення, зокрема дають можливість особам з вадами слуху активно долучатись до комунікаційних процесів як між

собою, так із людьми, що не мають цих вад. З іншого боку, розроблення системи комп'ютерного перекладу української жестової мови на українську словесну та навпаки дасть змогу широкому загалу людей вільно спілкуватися з нечуючими людьми.

Застосування сучасних інформаційних технологій для надання інформації особам з вадами слуху є актуальною науково-прикладною проблемою, для вирішення якої активно шукаються нові підходи [1, 2, 4], створюються нові методи та засоби [5–9].

Жестова мова – самостійна, природна мова, що 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. Послідовність слів у реченні жестової мови відрізняється від словесної мови.

Українська жестова мова, як й інші відомі жестові мови, має свою граматику і правила, відмінні від української словесної мови. Ця мова незалежна від української словесної мови, і не слід її плутати з калькованою жестовою мовою, яка повністю відповідає УСМ.

Українська жестова мова на відміну від калькованої жестової мови, яка повністю повторює послідовність слів речення словесної мови, має свій синтаксис та правила побудови речення. Переклад з української жестової мови на українську словесну є складною науково-прикладною проблемою, вирішення якої потребує аналізу граматики УЖМ, побудови правил перекладу з української словесної мови на жестову та навпаки.

Незважаючи на відсутність прийменників та відмінків, українська жестова мова є багаторівневою лінгвістичною системою, що володіє широким набором лексичних, граматичних засобів для висловлення думки та аналізу інформації.

Складність розроблення комп'ютерної системи перекладу української жестової мови посилюється відсутністю великих словників та корпусів УЖМ [3]. Оскільки жестові мови використовують для подання інформації жести та вирази обличчя замість звуків, то для адаптування систем перекладу писемних мов для

перекладу жестової мови необхідно розробити систему запису елементів жестової мови, створити відповідні перекладацькі словники й системи граматичних правил для розбирання речень жестової мови. Крім того, суттєва лінгвістична відмінність жестової мови від словесної ускладнює переклад.

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 (T. 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-порограмні 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 дисертаційному дослідженні для анотування української жестової мови обрано систему запису жестів у вигляді глосів, оскільки ця система є доволі зручною для комп'ютерного подання жестів та часто використовується у системах комп'ютерного перекладу іноземних жестових мов.

Проблему перекладу іноземних жестових мов досліджували С. Моррісей та Е. Вей (S. Morrissey and A. Way), Д. Стайн, Ф. Дрю, Г. Ней (D. Stein, P. Dreuw, H. Ney) (для англійської жестової мови), Р. Сан-Сегундо, А. Перез, Д. Ортіз, Л. Д. Д'Наро, М. І. Торез, Ф. Касакуберта (R. San-Segundo, A. Pérez, D. Ortiz, L. F. D'Haro, M. I. Torres, F. Casacuberta) (для іспанської жестової мови), Я. Банджерота (J. Bungeroth) (для німецької жестової мови).

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

У дисертаційному дослідженні розроблено системи комп'ютерного перекладу анотованої української жестової мови на основі граматично доповненої онтології, яка не має прямих аналогів у відомих наукових

дослідженнях. Створені під час дисертаційного дослідження методи та програмні засоби опрацювання та перекладу української жестової мови можна використовувати під час розроблення програм перекладу з однієї мови на іншу, що, своєю чергою, може полегшити спілкування осіб із вадами слуху між собою і з тими, хто не володіє жестовою мовою.

Зв'язок роботи з науковими програмами, планами, темами

Дисертаційні дослідження виконувались відповідно до напрямів наукової діяльності кафедри інформаційних систем та мереж Національного університету «Львівська політехніка», державної науково-дослідної роботи «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 граматичну онтологію для опису та перекладу української жестової мови, систематизувала та класифікувала правила перекладу для української жестової мови, створила систему комп'ютерного перекладу з української словесної мови на 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 виконано такі завдання:

- 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

Наукові дослідження виконано з метою їх практичного втілення реалізації та впровадження. Результати досліджень стали основою для створення програмно-алгоритмічного комплексу перекладу української жестової мови.

Практичне значення результатів роботи зумовлено тим, що вони дають змогу здійснювати комп'ютерний переклад анотованої української жестової мови на українську словесну мову та навпаки, що полегшує спілкування осіб із вадами слуху між собою і з тими, хто не володіє жестовою мовою.

Цінними для практики є такі результати дисертаційного дослідження:

- розроблено граматично доповнену онтологію УЖМ та УСМ;
- розроблено методи семантично-синтаксичного аналізу української словесної мови та анотованої української жестової мови;
- розроблено систему правил перекладу з української словесної мови на анотовану українську жестову мову;

- розроблено предметно-орієнтовану мову опису граматично доповненої онтології.

Результати дисертаційного дослідження використані для підготовки та створення навчальних матеріалів, поданих анотованою українською жестовою мовою у Львівській спеціальній загальноосвітній школі-інтернаті Марії Покрови для глухих дітей, для комунікації людей з вадами слуху у медичному центрі “Біокурс”, а також використані у навчальному процесі Національного університету “Львівська політехніка” при читанні курсу лекцій з дисциплін “Математична лінгвістика”, “Мови опису онтологій” та “Організація баз даних та знань”, що підтверджено відповідними актами.

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 MEMS» (с. Поляна, Україна, 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 ІТ-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).

Результати дисертаційних досліджень неодноразово апробовані на наукових семінарах кафедри інформаційних систем та мереж.

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 та була відзначена дипломом п'ятої Національної виставки «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 видань України, 8 публікацій у матеріалах і тезах конференцій, 1 наукове видання-брошура (препринт).

РОЗДІЛ 1

ОГЛЯД ВІДОМИХ ДОСЛІДЖЕНЬ ЖЕСТОВОЇ МОВИ

1.1. Відомі методи та засоби перекладу жестових мов

До першої половини ХХ ст. граматична структура жестових мов залишалася маловивченою та недостатньо формально описаною. Лінгвістичне дослідження жестових мов розпочалося в 50-х роках ХХ ст. у США, а першими дослідниками жестових мов були Бернард Тервурт (Bernard Tervoort) [21] та Вільям Стокоу (William Stokoe). Саме В. Стокоу вперше у книзі «Sign language structure» («Структура жестової мови»), виданій у 1960 р., описав морфологію жестової мови [22].

У США визнання жестової мови як окремого об'єкта наукового лінгвістичного дослідження відбулось лише у 70-х роках ХХ ст., а в Європі – після 80-х років.

Лінгвістичні дослідження польської жестової мови розпочав у 1994 р. вчений М. А. Ферріс. У його роботах описано структуру польської жестової мови [23]. У 1999 р. при Варшавському університеті створено дослідницьку групу під керівництвом М. Свіджинського, яка займається розробленням граматики польської жестової мови.

У Росії жестову мову вивчали Р. М. Боскіс [24] і Н. Г. Морозова [25], які зробили висновок, що у жестової мови є своя граматика, відмінна від граматики російської мови. Г. Л. Зайцева [26] описала склад дактильного алфавіту, лексику і граматику російської жестової мови.

Вивчення лінгвістики жестової мови в Україні розпочав Р. Г. Краєвський [27]. Вчений здійснив лінгвістичний опис мови жестів на основі українознавчого матеріалу і створив унікальний жестовий словник у вигляді посібника «Мова жестів глухих» 1964 р. (усього описано понад 1500 жестів), де лексеми

групується за конфігурацією кисті руки. Для кожного жесту описано просторове положення і спосіб руху кистей рук. Сьогодні в Україні вивченням синтаксичних особливостей УЖМ займаються Н. Б. Адамюк, І. І. Чепчина [28], С. В. Кульбіда [29].

Постійні зміни, що відбуваються в жестовій мові, призводять до труднощів перекладу. Через відсутність у жестовій мові чіткої системи, чуучим людям складно її вивчити, що природно зменшує кількість сурдоперекладачів. Для вивчення жестової мови та її відповідного структурування зазвичай необхідно докласти багато зусиль, затратити часові та людські ресурси. У низці країн Європи, а також у США та Канаді доволі багато зроблено щодо формування та зміцнення суспільного статусу жестової мови. Сайт Європейського союзу глухих повідомляє, що статус жестової мови на конституційному рівні зафіксовано у Фінляндії, Австрії, Чехії, Словаччині, Іспанії, на національному рівні жестова мова визнана у Великобританії, Італії, Франції, Греції, Німеччині, Швеції, Норвегії), невизначено статус жестової мови у Польщі, Румунії, Болгарії.

У світі не існує єдиної універсальної жестової мови. Жестові мови різних країн розвивалися незалежно, а тому значно відрізняються одна від одної. Крім того, кожна країна має місцеву жестову мову або різноманіття мов, що відображають культуру людей із вадами слуху, які там проживають. Наприклад, у Швейцарії існує чотири місцеві жестові мови, якими користуються глухі французької, італійської, німецької та швейцарської громад [30]. Попри використання єдиної словесної мови в Німеччині та Австрії, австрійська жестова мова (Austrian Sign Language, ÖGS) та німецька жестова мова (German Sign Language, DGS) є двома різними, не пов'язаними між собою мовами (окрім того, у Німеччині наявні декілька діалектів DGS). Подібна ситуація спостерігається й у США та Великобританії: словесною офіційною мовою цих країн є англійська. Однак люди із вадами слуху в Америці спілкуються американською жестовою мовою – амсленом (American Sign Language, ASL), а у Великобританії – британською жестовою мовою (British Sign Language, BSL) [31].

Для перекладу іноземних жестових мов розроблено низку програмно-алгоритмічних засобів, таких як: програма iCommunicator [32] (для англійської жестової мови), програма ViSiCAST [33] (для англійської жестової мови), програма TEAM Project [34] (для англійської жестової мови), програма Artemis [35] (для французької жестової мови), проект SASL-MT [36] (для південноафриканської жестової мови), програма THETOS [37] (для польської жестової мови). Проте зазначені програмно-алгоритмічні засоби не можна використати для перекладу УЖМ.

Програма iCommunicator реалізує такі функції:

- переклад усного речення на текст;
- переклад усного або письмового речення жестовою мовою;
- озвучення тексту речення;
- подання тексту і перекладу жестовою мовою наперед підготовленої лекції для студентів, які нечують.

Компанія Vcom3D (Флоріда, США) розробила низку програмних продуктів для американської жестової мови: Sign Smith Studio, Ілюстрований словник Sign Smith, ASL Animations, Gesture Builder [38].

Sign Smith Studio – це програмний комплекс, призначений для додавання користувачами інформації жестовою мовою з використанням анімованого аватара до їхнього цифрового контенту. Створений контент експортується у відео, яке потім можна додати до мультимедійних програм, веб-сторінок, презентацій PowerPoint. У базі даних міститься 2 000 жестів і виразів обличчя, 10 000 англійських слів і відповідних їм жестів. Інструменти редагування дозволяють користувачеві переміщати, видаляти, редагувати, вставляти і змінювати інформацію самостійно. Sign Smith Studio надає можливість разом з Gesture Builder створювати нові та спеціалізовані жести.

Ілюстрований словник Sign Smith призначений для тих, хто хоче вивчити основи американської жестової мови (American Sign Language, ASL). Ілюстрований словник містить 500 жестів, що найчастіше включаються до

початкових курсів ASL. Кожному слову в словнику відповідає картинка, використання цього слова в реченні та відображення за допомогою вибраного аватара. Щоб перейти до слів потрібно натиснути клавішу із зображення букви, з якої починаються слово або вибрати одну з 20 об'єктних категорій, таких як діяльність, прикметники, тварини тощо. Якщо слово має декілька значень, користувач вибирає слово, яке найбільше підходить по змісту. Слово, для якого в системі немає жестового відповідника, буде демонструватись побуквенно.

ASL Animations – єдиний сьогодні анімований ресурс для вивчення жестової мови, який швидко і зручно інтегрується у програмне забезпечення. Призначені для допомоги користувачам у вивченні значення слів і встановлення відповідності між англійським текстом і американською жестовою мовою. 5 томів ASL Animations містять більш ніж 1100 жестів, що подають більш ніж 1800 англійських слів.

Вищезгадані комп'ютерні програмно-алгоритмічні засоби для перекладу жестової мови є платними.

Розроблена дослідницькою командою IBM Hursley програма SiSi (Say It Sign It) (Великобританія) [39] автоматично перетворює речення англійської словесної мови на речення британської жестової мови (BSL), після чого речення відтворюється за допомогою аватара. Програма SiSi поєднала в собі низку оригінальних комп'ютерних технологій. Модуль розпізнавання речень англійської словесної мови перетворює кожне слово речення на текст, який у подальшому інтерпретується в жести, що використовуються для анімації аватара.

«SignSpeak» [40] один із перших проектів, профінансованих ЄС, спрямований на вирішення проблеми автоматичного розпізнавання та перекладу жестової мови. Загальною метою проекту «SignSpeak» є надання доступу до інформаційних ресурсів особам з вадами слуху. Проект «SignSpeak» поєднує інноваційні наукові підходи та передові технології для аналізу зображень,

автоматичного розпізнавання мови та статистичного машинного перекладу в межах інтегрованої структури.

Програмний комплекс «SLEd» (Sign Language Education Program, російсько-англійсько-російський анімований словник жестової мови) розроблений за підтримки Інституту «Відкрите суспільство» (Фонд Сороса), Росія.

База даних SLEd містить порівняно невеликий набір жестів російської та американської жестових мов, обмежений тематикою і фразеологією шкільної програми вивчення англійської мови. Програмний комплекс функціонує в середовищі Windows 95/98 і не потребує значних апаратних ресурсів. Програмний застосунок можна вважати прототипом програмно-алгоритмічного комплексу, що його розробляють російські вчені в межах проекту «Фонд жестової мови» (автор проекту О. Л. Воскресенський) для перекладу зі словесної мови (як базова мова використовується російська мова) на жестову мову, а також для перекладу з однієї жестової мови на іншу [41].

Анімований словник російської жестової мови призначений для використання в мережі Інтернет, в ньому передбачені інтерактивні інтелектуальні засоби формування вказівок віддаленим користувачам щодо конкретнішої модифікації жесту. Думки користувачів підсумовуються з метою визначення найпопулярнішої модифікації.

Для італійської жестової мови розроблено словник LIS (Italian Sign Language) [42], який надає можливість користувачеві отримати інформацію про жести та слова, що відповідають певним жестам. Цей словник розроблений для використання у мобільних телефонах зі всіма відомими типами мобільних операційних систем. Словник LIS забезпечує інтеграцію з базою даних MultiWordNet, що дає змогу використовувати значення слів у певному контексті, записаних у MultiWordNet. Словник LIS містить більш ніж три тисячі жестів, для анімації яких використовується аватар. Кожен жест пов'язаний з відповідною італійською лемою та набором синсетів.

Такі інформаційно-технологічні розроблення, крім комп'ютерного застосування, мають велике соціальне значення, зокрема дають можливість особам з вадами слуху активно комунікувати між собою та спілкуватись із людьми без цих вад. Розроблення та комп'ютерна реалізація української мови жестів дасть змогу широкому суспільному загалу вільно спілкуватися з нечуючими людьми.

Для перекладу на жестову мову використовуються методи перекладу на основі прикладів [43], методи перекладу на основі правил [44, 45], методи перекладу на основі даних [46], методи перекладу на основі фраз [47] та статистичні методи перекладу [48, 49]. Іспанські дослідники [44] порівняли підходи на основі правил, на основі фраз та статистичного підходу для перекладу на іспанську жестову мову. За наведеними даними [44] найточнішим виявився підхід на основі правил. Проте автору не відомі роботи, в яких досліджувалась застосовність наведених методів для перекладу УЖМ (табл. 1.1).

Я. Банджерот та Г. Ней запропонували застосовувати статистичний підхід до перекладу німецької жестової мови у текст та навпаки [49]. Переклад складається з таких етапів: розпізнавання вхідних жестів та запис розпізнаних жестів у вигляді глосів, статистичний переклад та запис вихідного тексту. Для тестування і навчання системи перекладу вибрано корпус, який містить 1399 речень німецькою жестовою мовою (German Sign Language, DGS). Оцінка результату перекладу здійснюється за допомогою оцінки помилок перекладу слів WER (Word Error Rate) та позиційно-незалежної оцінки помилок PER (Position-Independent Word Error Rate) та становить відповідно 85,4 % і 43,9 %.

У роботі [48] розглянуто модель статистичного машинного перекладу анотованих за допомогою глосів речень американської жестової мови у речення англійської словесної мови. Складовими перекладу є створення та вирівнювання корпусу паралельних текстів, розроблення декодера для статистичного машинного перекладу із залученням мовних моделей (англійської мови та американської жестової мови), покращення результатів перекладу.

Таблиця 1. 1

Методи машинного перекладу жестових мов

Методи машинного перекладу:	У світі	В Україні
Методи на основі правил	Р. Сан-Сегундо, А. Перез, Л. Д. Д'Гаро, М.І. Торез, Д. Ортіз, Ф. Касакуберта (2010)	–
Методи на основі прикладів	С. Моррісей (2005)	–
Методи на основі даних	Н. Строппа та Е. Вей (2006)	–
Статистичні методи	Д. Стайн, Я. Банджерот, Г. Ней (2006), Ф. Дрю (2007)	–
Методи на основі граматичних конструкцій речень	–	Ю. В. Крак, С. О. Романишин, О. В. Бармак (2011)
Методи з використанням онтології	А. М. Алмасуд, Х. С. Аль-Халіф (2012)	–
Методи з використанням граматично доповненої онтології	–	–

Схему функціонування системи статистичного машинного перекладу подано на рис. 1.1. Для тестування системи статистичного машинного перекладу використовували корпус RWTH-Бостон-104. Під час перекладу речень

американської ЖМ на речення англійської словесної мови оцінка помилок перекладу WER становить 21,2 %, а позиційно-незалежна оцінка помилок перекладу PER – 20,1 %.

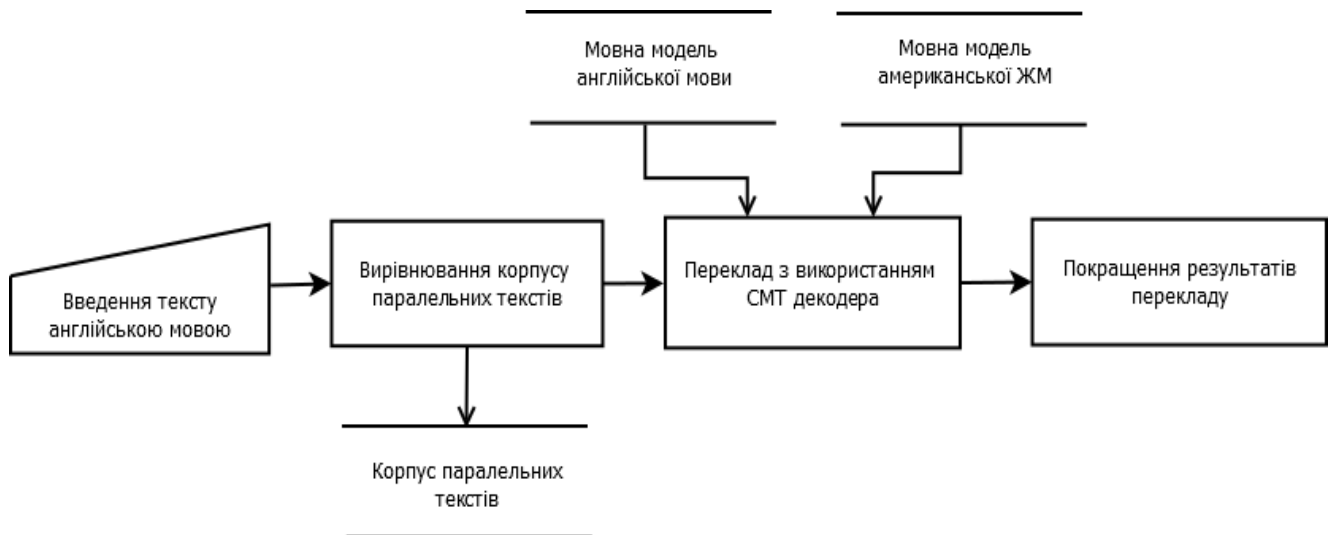


Рис. 1.1. Схема функціонування системи статистичного перекладу речень американської жестової мови у речення англійської словесної мови

Індійські вчені Т. Дасгупта та ін. [45] запропонували модель машинного перекладу на основі правил для трансформації речень індійської словесної мови на речення індійської жестової мови. Переклад на жестову мову складається з п'яти етапів: парсеру вхідного тексту, побудови ЛФГ-структури (лексико-функціональна граматики), трансформації вхідного речення у вихідне з використанням правил граматики, генерування жестів, синтезу індійської жестової мови. Загальна оцінка якості системи перекладу становить близько 90 %.

У роботі іспанських вчених Р. Сан-Сегундо та ін. [44] використано підхід на основі правил для перекладу речень іспанської словесної мови у речення іспанської жестової мови. Переклад реалізують за допомогою таких етапів: розпізнавання мовлення та його перетворення на послідовність слів, трансформація послідовності слів у послідовність жестів з використанням правил граматики, анімація жестів. Переклад здійснюється з використанням бази

правил граматики іспанської жестової та словесної мов. Систему перекладу на основі правил для іспанської жестової мови зображено на рис. 1.2. Результат функціонування системи перекладу на основі правил поданий за допомогою таких оцінок: автоматичної оцінки якості перекладу BLEU (Bilingual Evaluation Understudy) становить 0,72 %, оцінки помилок перекладу жестів SER (Sign Error Rate) – 16,75 % та позиційно-незалежної оцінки помилок перекладу PER – 13,17 %.

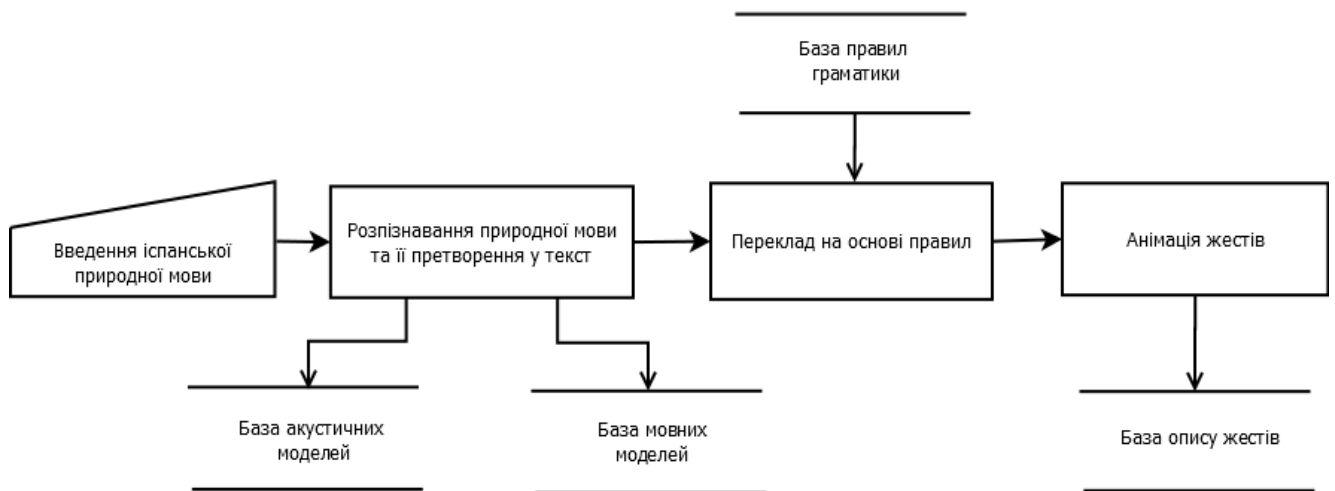


Рис. 1.2. Схема функціонування системи перекладу на основі правил для іспанської жестової мови

Інформаційна технологія перекладу польської жестової мови подана у роботі [37]. Для польської жестової мови розроблено систему перекладу на основі правил, яка доповнена такими процесами, як синтаксичний та семантичний аналіз вхідного тексту, перетворення структури «предикат–аргумент» вхідного речення польською мовою на структуру «предикат–аргумент» вихідного речення жестовою польською мовою та анімація відповідних жестів. Семантичний аналіз речень здійснюється з використанням словника Semsyn та бази правил побудови структури «предикат–аргумент». Схему функціонування системи перекладу речень польської словесної мови у речення польської жестової мови наведено на рис. 1.3. Для цієї системи перекладу автор не наводить оцінки результатів перекладу.

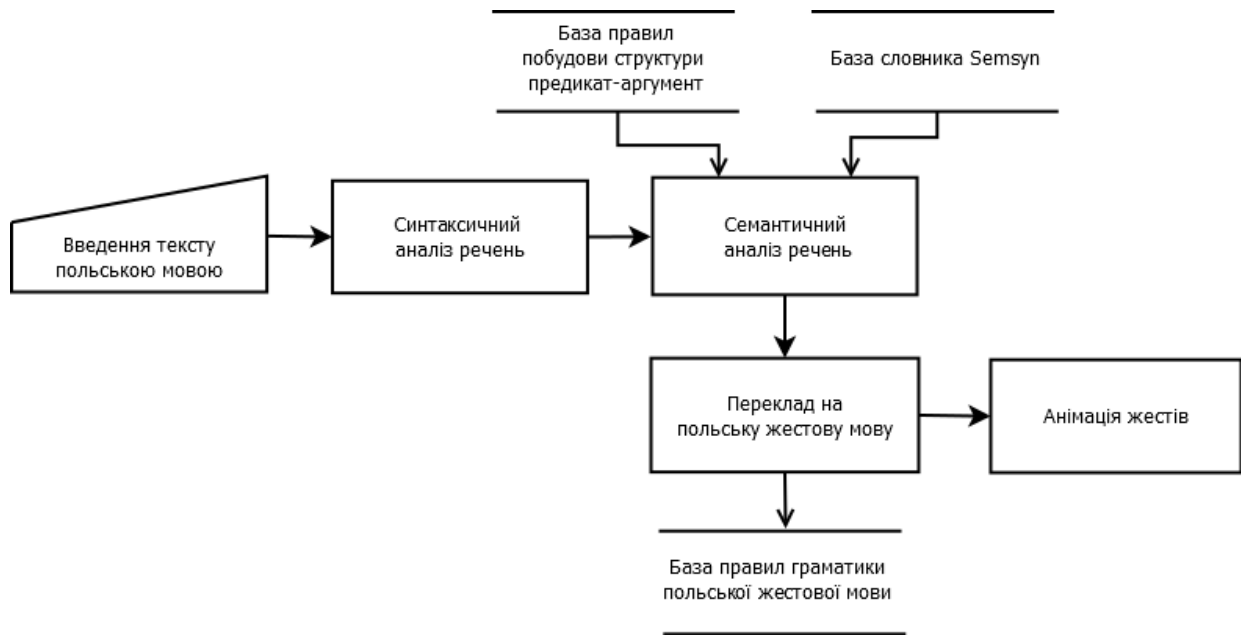


Рис. 1.3. Схема функціонування системи перекладу на основі правил для польської жестової мови

Статистичні системи машинного перекладу [48, 49] та системи перекладу на основі правил [44, 45] не використовують онтології предметних галузей.

Для перекладу арабського тексту на арабську жестову мову вчені А. Алмасуд та Х. О.-Халіф [50] запропонували модель перекладу на основі онтологій арабської словесної та арабської жестової мов. Переклад складається з таких етапів: морфологічний аналіз вхідного арабського тексту, трансформація тексту з використанням правил граматики, переклад на основі онтологій арабської словесної та арабської жестової мов (рис. 1.4). Трансформація речень відбувається із залученням бази правил граматики арабської жестової мови. Результати перекладу у відсотковому еквіваленті для цієї системи її автори не наводять.

Українські вчені Ю. В. Крак, О. В. Бармак та ін. з Інституту кібернетики ім. В. М. Глушкова НАН України проводять значні наукові дослідженнями УЖМ [51]. Вчені запропонували об'єднувати речення в узагальнені структури, які міститимуть замість слів їхні послідовності, кожен з яких можна використовувати для побудови речення. Вони побудували 764 такі узагальнені

граматичні конструкції із 6559 речень, а також виділили узагальнені конструкції, які використовують не фіксовані множини слів, а ознаки граматичних категорій для групування. Українські вчені запропонували та реалізували інформаційну технологію автоматизованого перекладу з флективних мов на жестову з використанням інфологічних моделей словників. Ця технологія складається з таких процесів: визначення параметрів словозміни для кожного слова у вхідному реченні, пошуку граматичної конструкції вхідного речення, пошуку відповідної граматичної конструкції жестового речення, заміни слів на відповідні жести.

Для тестування запропонованої технології дослідники розробили відповідне програмне забезпечення. За допомогою цієї технології експериментально отримано однозначний переклад без спотворення змісту в 100 % речень, узятих із шкільної програми вивчення української мови жестів. Проте для досягнення якісного комп'ютерного перекладу української жестової мови важливим є застосування методів семантичного аналізу.

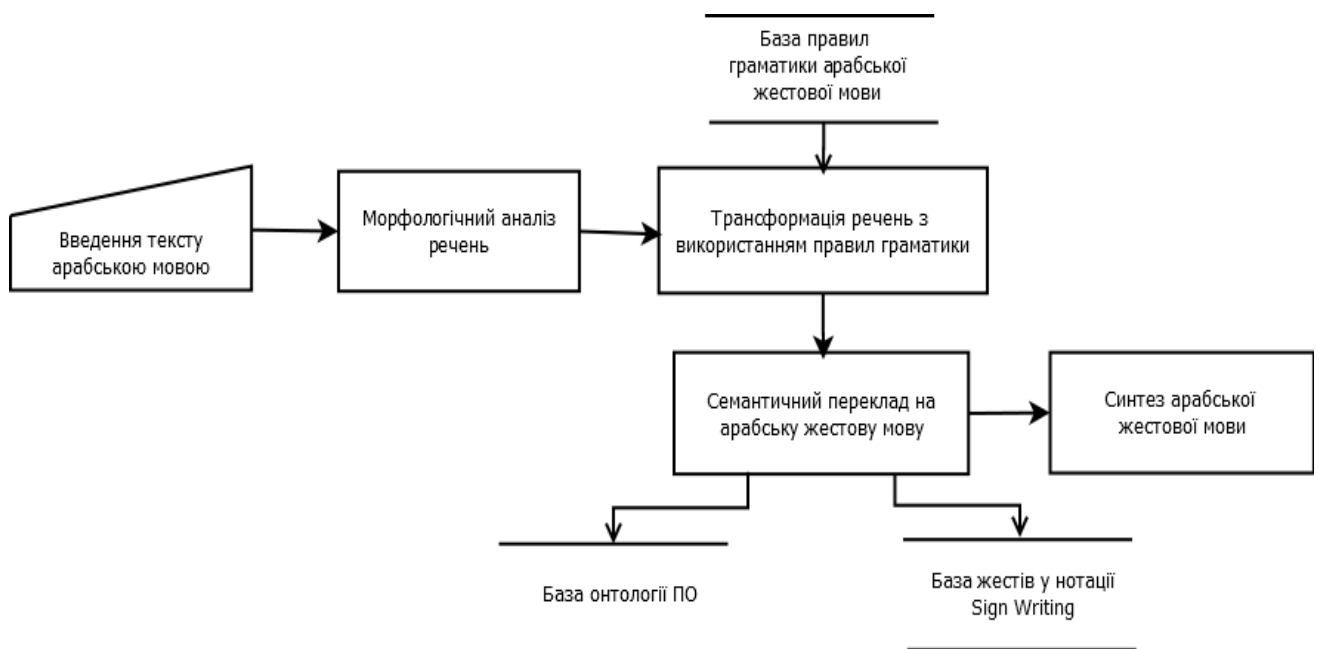


Рис. 1.4. Схема функціонування системи перекладу на арабську жестову мову з використанням онтології

Незважаючи на те, що лінгвістичні дослідження української жестової мови науковці проводять вже тривалий час, досі немає повного опису граматики УЖМ та не сформовані базові системні принципи перекладу УЖМ.

1.2. Методи запису жестової мови

Жестова мова – це візуально-просторова мова, і тому не може бути записана аналогічно до того, як це робиться у будь-якій словесній мові. Для практичного, зокрема комп’ютерного, застосування у світі розроблені кілька систем запису жестової мови (нотації), які мають ознаки наочності та уніфікованості, що є визначальним для швидкого введення жестів у словник.

Першу систему запису жестової мови розробив У. Стокоу [22]. Вона складалася з 55 символів, 12 з яких описують місце виконання жесту, 19 простих символів для позначення конфігурацій елемента, що діє, (рук, або руки) та 24 символи для зображення жестової дії (руху / виконання жесту). Для позначення форми руки У. Стокоу використовував цифри (наприклад, символ 5 позначав розкриту долоню з розчепіреними пальцями) і букви англійського алфавіту. Орієнтація руки позначається в нотації нижнім індексом при символі форми руки. Порядок символів під час запису жесту фіксований: спочатку вказувалося місце виконання жесту, потім форма руки, наприкінці особливості руху руки.

Нотацію У. Стокоу (рис. 1.5) використовували для запису першого словника американської жестової мови [52].

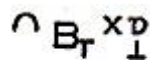


Рис. 1.5. Запис виразу «не знаю» у нотації У. Стокоу

Одним із суттєвих недоліків нотації У. Стокоу є те, що вона не дає змоги фіксувати вираз обличчя, який є невід’ємною складовою багатьох жестів. Також за допомогою цієї нотації не можна описати такі параметри виконання жесту, як амплітуду руху, різкість/плавність жесту, які, як показали подальші дослідження жестових мов, мають велике значення для подання жесту [53].

Однією з найвідоміших систем нотації, значно поширеної на Американському континенті, є нотація SignWriting [54], яку розробив В. Саттон. Проте вона містить доволі великий набір знаків (понад тисячу), що робить її малопридатною для застосування в комп'ютерних технологіях навіть з використанням можливостей кодування Юнікод. У SignWriting жест зображується піктограмою (рис. 1.6), що містить іконічні зображення рук, їхніх рухів (за допомогою різноманітних стрілок), та їхнє розташування відносно один одного і мовця, за необхідності передається вираз обличчя або напрям погляду, супутні жести. Ланцюжок жестів записується згори донизу. Транскрипція дає змогу точно і наочно відображати жести, проте залишається простою для сприйняття.

Більшість носіїв жестової мови використовують систему SignWriting, вважаючи її письмовою формою жестової мови [53]. SignWriting активно використовують під час створення словників жестових мов.



Рис. 1.6. Запис у системі нотації SignWriting слова «глухий»

Система нотації жестів, яку запропонувала Л. Дімскіс [55], містить 30 конфігурацій рук, близько 50 характеристик місця виконання жесту і більш ніж 70 характеристик локалізації (рис. 1.7).



Рис. 1.7. Приклад запису числа «чотири» в нотації Л. Дімскіс

Проте найбільше поширення в світі останніми роками отримала гамбурзька система нотації, запропонована Т. Ханке, більш відома фахівцям під

скороченою назвою HamNoSys (від англ. Hamburg Notation System) [56], розроблена наприкінці XX ст. у Німеччині. Ця система нотації може застосовуватися для запису доволі великої кількості жестових мов: усі параметри форми жесту відображуються в ній за допомогою іконічних значків (на відміну від транскрипції У. Стокоу, де символи для форми руки не є іконічними). HamNoSys орієнтована на дуже докладний опис жесту. У ній використовується близько двохсот символів, позначаються не тільки місце виконання жесту, форма і траєкторія руху руки, але й орієнтація руки, немануальні жести (міміка мовця). Під час записування жесту спочатку описується форма руки (рук), орієнтація, місце виконання жесту й особливості руху. Жести в транскрипції HamNoSys доволі складні для візуального сприйняття, проте ця система запису є зручною для комп'ютерного опрацювання (рис. 1.8).

Транскрипція HamNoSys використана, зокрема, в тематичних словниках німецької жестової мови [57], словнику австралійської жестової мови [58], новозеландської жестової мови [59] та інших.



Рис. 1.8. Приклад запису числа «дев'ятнадцять» в нотації HamNoSys

Відомою є і система позначення жестів за допомогою глосів. Глоси – це текстове подання жестів (жести записуються великими літерами), яке часто використовується для розшифрування жестової відеопослідовності [48]. Основним недоліком цього подання є те, що глоси не передають вираз обличчя та інформацію про рух тіла [50]. Це, своєю чергою, не дає змоги показувати відмінність між різними напрямками рухів рук, за допомогою яких глухі асоціюють різні об'єкти у просторі.

1.3. Подання словників жестової мови в комп'ютеризованих системах

Для систем машинного перекладу особливе значення мають корпуси паралельних текстів, тобто оригінальних текстів та їх перекладів, об'єднання яких у комп'ютерній програмі опрацювання паралельних текстів дає змогу для певного сегмента оригіналу автоматично встановлювати відповідний йому сегмент у перекладі [60].

1.3.1. Корпуси паралельних текстів

Майже всі загальнодоступні ресурси для перекладу жестової мови, які записані в лабораторних умовах у дослідницьких цілях, мають загалом невеликий розмір словника, а також тісно пов'язані з певними правилами запису. Системи перекладу на основі даних, які автоматично навчались на цих корпусах текстів, не завжди генерували правильні вихідні речення, тому що структура вхідних речень часто була розроблена заздалегідь [61], або мала невеликі варіації [40, 49] у результаті чого потрібно було застосовувати більш придатні моделі мови. Крім того, більшість самостійно записаних корпусів складається з обмеженої кількості жестів [61, 63].

У системах машинного перекладу жестової мови для експериментів зазвичай використовуються корпуси паралельних текстів з певної предметної галузі. Наприклад, для перекладу іспанської жестової мови [44] використовується корпус, що складається з 416 пар речень, які містять 650 слів та 320 жестів. Для того, щоб записати іспанську жестову мову (LSE) з погляду текстових символів, кожен жест записано словом, поданим великими літерами. Цей корпус містить речення з предметної галузі «Паспортний стіл» та отримані з діалогів між офісними працівниками та людьми, котрі хочуть отримати паспорт.

У системі перекладу на жестову мову [64] використано корпус ATIS (Air Travel Information System), що містить дані з предметної галузі авіаподорожей. Корпус складається з 600 речень для англійської та німецької словесної мов, ірландської, німецької та південноафриканської жестових мов. Завдяки наявності мовних пар, є можливість прямого перекладу між жестовими мовами.

Для перекладу німецької жестової мови [65] використовується корпус RWTH-PHOENIX-v3.0, що містить дані зі сфери новин, зокрема розділу – прогноз погоди. Корпус складається з 3077 пар речень німецькою жестовою та словесною мовами. Речення жестовою мовою анотовані у вигляді глосів.

Для тестування системи машинного перекладу, описаної у роботі [66], вибрано корпус RWTH-Бостон-104. Цей корпус сформований у Бостонському університеті для проведення лінгвістичних досліджень та містить 201 відеоречення, які складаються зі 104 неповторювані слова. Речення демонструють три користувачі жестової мови, а корпус розбитий на навчальну (161 речення) та тестову (40 речень) множини. Структура речень проста, тому рівень складності моделі не високий.

Для італійської жестової мови сформовано корпус AEWLIS [67], який містить 585 речень. Корпус AEWLIS для кожного жесту в реченні містить таку інформацію: запис жесту у вигляді глоса, однина/множина жесту, напрямок жесту, виконання жесту лівою, правою чи обома руками, положення тіла тощо, також вказується роль жесту та тип речення (розповідне, питальне чи заперечне).

Корпус паралельних речень іспанської словесної та жестової мов описаний у роботі [44]. Цей корпус містить 4080 пар речень та базу даних жестів, записаних у декількох нотаціях: за допомогою глосів, у нотації HamNoSys та алфавітній нотації SEA (Sistema Escritura Alfabética). База даних жестів містить жести з паралельного корпусу і жести для позначення літер (необхідних для написання слова), цифр від 0 до 100, жести для специфікації часу, місяця та днів тижня.

У системі перекладу на основі фраз із чеської жестової мови на чеську та навпаки [46] описано експерименти, які проводились на корпусі паралельних текстів чеської жестової та словесної мов. Корпус містить 1130 діалогів, взятих з телефонних розмов між клієнтами та операторами інформаційного центру розкладу руху поїздів. Проведено порівняння якості машинного перекладу з ручним.

Українськими вченими під керівництвом Ю. В. Крака [68] розроблено інфологічну модель бази даних для узагальнених граматичних конструкцій УСМ та УЖМ у вигляді таблиць та зв'язків між ними. У роботі [51] запропоновано практичну реалізацію інфологічних моделей для словників флективної та жестової мов для роботи з даними граматичними конструкціями і алгоритм перекладу. Описано інфологічні подання словників української мови як множини інфінітивів, флексій та параметрів словозміни слів і жестової мови як множини жестів та параметрів, що їх характеризують. Отримано експериментальне підтвердження ефективності викладеного підходу на базі української мови.

Проте, у роботах [51, 68] не враховано те, що кожна граматична конструкція має певне семантичне навантаження. Наприклад, у граматичних конструкціях «я йду» та «час йде» жест «ЙТИ» демонструється по-різному. Тому ми не можемо цей жест перекласти однозначно без урахування контексту (семантичного навантаження). Для вирішення цього завдання потрібно створити словник концептів та їхніх відповідників в УСМ та УЖМ.

1.3.2. Онтологічні словники

Тривалий час базовим джерелом знань для систем перекладу були тлумачні та двомовні словники [69]. Через те, що тлумачні словники не наділені необхідною повнотою і структурою для вираження семантичних зв'язків між словами і поняттями, замість них почали використовувати онтологічні словники [70]. 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.

У сучасних інформаційно-технологічних розробленнях для структурування знань із різних предметних галузей використовуються онтології. Цей підхід продемонстрував свою ефективність для формальної специфікації

понять і відношень між ними, які характеризують певну галузь знань. Перевагою онтологій як способу представлення знань є їх чітка формальна структура, що, своєю чергою, суттєво спрощує реалізацію процедур комп'ютерного опрацювання [71].

Відомі роботи з розроблення як спеціалізованих онтологій для опису окремих предметних галузей, так і загальних, для опису загальноживаних понять мови. Наприклад, розроблені онтології у галузі медицини (Gene Ontology для анотації біомедичних даних [72], система уніфікованої медичної мови (UMLS, the Unified Medical Language System) [73], онтологія у галузі товарів та послуг UNSPSC [74] та ін.

Загальні онтології широко використовуються для досліджень в галузі комп'ютерної лінгвістики. До таких онтологій належать WordNet [75] та ConceptNet [76] для англійської мови, plWordNet для польської мови [77], CWN (Chinese Wordnet) для китайської мови [78], WOLF (WordNet Libre du Français) для французької мови [79], MultiWordNet project для італійської мови [80], BalkaNet project для шести європейських мов (болгарської, чеської, грецької, румунської, турецької та сербської) [81], GermaNet для німецької мови [82], IndoWordNet для індійської мови [83], проект RussNet [84] для російської мови.

Для української мови розробленням WordNet-подібного словника для частини іменників займається, зокрема, І. М. Кульчицький [85]. У результаті дослідження він розробив фрагмент WordNet-подібного словника української мови, в якому реалізовано 194 синсети, пов'язаних зв'язками гіпо-/гіперонімії, антонімії, а також додатково зв'язками меронімії/голонімії.

Над створенням української онтологічної лексико-семантичної бази знань UkrWordNet (UWN) працюють науковці Київського національного університету імені Тараса Шевченка [86]. Проект триває декілька років і в його межах досягнуто значних успіхів щодо мовної локалізації лексичних вузлів мережі, побудови та наповнення власних структур бази UWN, створення інструментарію

для автоматизації заповнення структур бази знань. На цей час ця онтологія містить близько 80000 концептів.

Сьогодні WordNet є найбільшим за обсягом програмно реалізованим лексиконом [85]. Базовою структурною одиницею WordNet є синонімічний ряд – синсет, який об'єднує лексеми з подібним значенням. Кожен синсет є деяким значенням, поняттям мови. Синсет можна розглядати як множину значень слів, які виражають одне поняття. Для кожного синсета визначаються приклади вживання лексем синсета у деякому контексті. Кожне слово однозначно ідентифікує один синсет. Між синсетами встановлені асоціативні відношення типу: гіпонім, гіперонім, синонім, голонім, меронім.

Для редагування онтологічних словників використовують два основні підходи – на основі бази даних і на основі текстових файлів, які потім перетворюються (компілюються) на базу даних. Підхід з базою даних потребує розроблення спеціальних програм для управління базою, синхронізації різних баз тощо. Підхід на основі файлів – засобів перевірки граматики і компіляції у базу даних.

Загально відома лексична онтологія WordNet складається з файлів, коду для перетворення цих файлів на базу даних, інтерфейсів для пошуку та відображення інформації з бази даних. Файли містять інформацію про іменники, дієслова, прикметники і прислівники та їх організацію у групи синонімів, а також описують відношення між даними групами [75].

Для перегляду і редагування словника Russian WordNet використовують редактор TenDrow, що дає змогу переглядати і редагувати словникові статті WordNet та ієрархії їхніх відношень. Як показала практика розроблення і побудови онтологічного словника, найефективнішим є редагування спеціально підготовлених текстових файлів і набір утиліт для внесення змін у базу даних. Для кожного етапу побудови словника формується набір таких файлів і засобів їх опрацювання. Редактор TenDrow призначений для створення і редагування широкого класу тезаурусів і близьких до них структур [87].

Українська онтологічна лексико-семантична база знань UkrWordNet використовує характерні для WordNet структурні елементи: синсети та набори семантичних (22 типи) і лексичних (12 типів) зв'язків. Проте онтологія UWN суттєво відрізняється від онтологій типу WordNet. З погляду архітектури UWN належить до онтологій створених на базі СУБД.

Наукових робіт, у яких використовують онтології для перекладу жестової мови, доволі мало. Зокрема, у роботі [50] описана система машинного перекладу арабської жестової мови, в якій використовується онтологія домену для предметної галузі «Релігія» та онтологія Arabic WordNet.

Проте онтологічні словники не описують повною мірою мовні конструкції для вираження семантичних зв'язків між поняттями, що зумовлює необхідність розроблення деталізованіших моделей онтологій, зокрема, це є актуальною проблемою для української словесної та української жестової мов.

1.4. Правила перекладу жестової мови

Під час перекладу словесної мови на жестову мову слід враховувати, що жести в багатьох випадках зіставляються з наборами слів, що виражають деяке поняття – концептами. Це призводить до того, що на виході системи перекладу з'являється певна кількість синонімічних значень (переданих одним і тим самим жестом або сукупністю жестів), деякі з яких синонімами з традиційного погляду не є, хоча і відображають у словесній мові ті самі значення (пор. «лампа висить під стелею» і «лампа прикріплена до стелі»).

Ще однією проблемою під час перекладу з жестової мови на словесну та навпаки є те, що в багатьох випадках той самий жест може означати іменник, дієслово, прикметник, і спроби розділити ці поняття якимись особливостями показу, як правило, не дають бажаних результатів.

Особливістю жестової мови є можливість передати практично весь спектр думок і понять значно меншою кількістю символів, аніж у словесній мові. Отже, жестова мова корінним чином відрізняється від словесної, і побудова речень у

ній теж зовсім інша. Крім того, існує значна відмінність у засобах передавання змісту речення (у жестовій мові велике значення відіграє порядок слів у реченні). Так само як і в словесній мові, в ЖМ прості речення мають певну класифікацію. За метою повідомлення вони поділяються на розповідні, питальні та спонукальні речення. *Розповідне* жестове речення містить повідомлення про якийсь факт, розповідь про щось. Розповідні речення поділяються, своєю чергою, на стверджувальні та заперечні речення. *Питальні* речення містять якесь запитання, покликане отримати інформацію, про щось дізнатися. *Спонукальне* речення виражає волевиявлення мовця, яке потребує виконання, а саме: наказ, прохання, пораду, застереження, протест, згоду [28]. У кожного речення є структурна основа, яка може зводитися до структури підмет-присудок («Іменна група [noun phrase] + Дієслівна група [verb phrase]») або до конструкції, що включає до свого складу необхідні члени речення («Підмет + Присудок + Доповнення»). Структурну основу двоскладного речення утворюють підмет і присудок. Двоскладні речення представлені в кожній мові кількома структурними схемами залежно від граматичного оформлення підмета й присудка і залежно від характеру зв'язку між ними.

Основною відмінністю жестових мов є відмінність у порядку слів у реченні. Послідовність слів у реченні жестової мови відрізняється від словесної мови. У жестовій мові немає закінчень, які у словесній мові передають відтінки думок, час, особу тощо. Тому передача думок часто залежить від супутніх жестів, особливо тих, що вказуються перед ним. Для передавання роду іменників використовуються жести «жінка» та «чоловік». Наприклад, кішку показуємо як жести «кіт» + «жінка». Число можна виражати двома способами. По-перше, можна додавати до певного жесту (вжитого на позначення однини) ще один жест, що означає «багато». По-друге, можна повторювати один жест кілька разів поспіль. Наприклад, для того, щоб показати людей, повторюємо кілька разів жест «людина». Відмінків у мові жестів немає [28]. Оскільки у жестовій мові відсутні прийменники, сполучники, частки, тому структура речення повинна

мати строгий порядок слів для уникнення нерозуміння. Незважаючи на відсутність прийменників та відмінків, жестова мова є багаторівневою лінгвістичною системою, що володіє широким набором лексичних, граматичних засобів для висловлення думки і аналізу інформації.

Для австрійської жестової мови стандартним порядком слів у реченні є Суб'єкт – Об'єкт – Предикат, а для американської жестової мови – Суб'єкт – Предикат – Об'єкт [88].

Суб'єкт – Предикат – Об'єкт є основним порядком слів для португальської, тайванської, японської [88] та хорватської жестових мов [89], хоча порядок Об'єкт – Суб'єкт – Предикат теж часто використовується.

Наприклад, індійській та аргентинській жестовим мовам характерний такий порядок слів: Суб'єкт – Об'єкт – Предикат [45], на відміну від англійської мови, в якій порядок слів є Суб'єкт – Предикат – Об'єкт. Наступний приклад показує зміну порядку слів під час перекладу з англійської словесної мови на індійську жестову:

Англійська словесна мова

«I have a computer»

Індійська жестова мова

«I COMPUTER HAVE».

Правила перекладу, які використовуються у системі машинного перекладу для індійської мови жестів, подані у роботі [45]. Ця система налічує 10 граматичних правил для синтезу індійської жестової мови з англійської словесної.

Іспанській жестовій мові характерний такий порядок слів: Суб'єкт – Предикат – Об'єкт [44]. У системі перекладу іспанської жестової мови використовують 170 правил перекладу, які розділяють на чотири типи: одне слово відповідає одному жесту, кілька слів утворюють унікальний жест, одному слову відповідає кілька жестів, четвертий вид правил – це правила, які генерують декілька жестів із декількох слів із врахуванням певних зв'язків між ними. Приклад ілюструє той факт, що порядок слів під час перекладу з іспанської словесної мови на іспанську жестову мову не змінюється (для зручності в

дужках вказано англійський переклад речень іспанської словесної та іспанської жестової мов):

Іспанська словесна мова

«Necesito mesa mi casa»

(«I need table for my house»)

Іспанська жестова мова

«NECESITAR MESA MI CASA»

(«I NEED TABLE MY HOUSE»)

Для перекладу китайської жестової мови розроблено спеціальні правила перекладу, оскільки китайська словесна та китайська жестова мови мають різну послідовність слів у реченні [90]. Наприклад, дієслова у реченні китайської жестової мови стоять після іменника, тоді як у реченнях китайської словесної мови іменник стоїть перед дієсловом. Загальні правила перекладу з китайської словесної мови на китайську жестову є такими:

- 1) правило перекладу іменника та прикметника: прикметник + іменник → іменник + прикметник;
- 2) правило перекладу предиката (дієслова) та іменника: дієслово + іменник → іменник + дієслово;
- 3) правило перекладу питальних речень: питальний займенник + іменник → іменник + питальний займенник;
- 4) правило пропущеного слова: слово, яке немає жестового відповідника → порожнє місце.

Висновки до першого розділу

Аналіз наявних систем перекладу жестових мов показав, що загалом переклад жестової мови є складною науково-технічною проблемою, вирішення якої потребує розв'язання низки задач, пов'язаних з усуненням неоднозначності слів, граматичним розбором жестових та словесних речень, збереженням та використанням знань експерта-перекладача тощо. Проведено огляд сучасних досліджень, які виконуються з метою побудови коректних систем комп'ютерного перекладу іноземних жестових мов.

Автор дослідив системи машинного перекладу жестових мов з використанням статистичних методів перекладу, методів перекладу на основі правил, методів перекладу на основі даних та методів перекладу на основі онтологій. Зазначені інформаційні технології не підходять повною мірою потребам та рівню задач, які виникають під час перекладу УЖМ.

Проаналізовано та проведено порівняння методів запису жестової мови, проаналізовано організацію словників жестової мови та онтологічних словників.

Аналіз функціональності відомих сучасних систем перекладу жестової мови на словесну дає змогу сформулювати низку першочергових задач, що виникають під час їх розроблення та функціонування. До списку невирішених наукових задач належать:

- формування системи позначень (нотація): адекватне письмове подання жестової мови;
- забезпечення гнучкості: створення системи для перекладу декількох мовних пар;
- врахування лінгвістичних особливостей: створення системи, яка може впоратися з лінгвістичними особливостями жестової мови;
- формування наборів даних: невелика кількість даних і лінгвістичних знань для жестових мов;
- забезпечення зрозумілості: інформація, перекладена на жестову мову системою перекладу, повинна бути доступною та зрозумілою для людей з вадами слуху, тобто відповідати реальній жестовій мові;
- виконання оцінювання систем перекладу: застосування методів оцінювання для порівняння та оцінки успішності та зручності використання кожної системи перекладу.

Провівши порівняльний аналіз відомих сучасних систем перекладу для жестових мов, можна зробити висновок про те, що найкращі та найточніші результати перекладу жестової мови спостерігаються під час використання процедур машинного перекладу, що ґрунтуються на правилах та на онтологіях.

РОЗДІЛ 2

ГРАМАТИЧНО ДОПОВНЕНА ОНТОЛОГІЯ ДЛЯ УКРАЇНСЬКОЇ СЛОВЕСНОЇ ТА УКРАЇНСЬКОЇ ЖЕСТОВОЇ МОВ

У другому розділі побудовано 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* опису для наповнення цієї граматично доповненої онтології.

2.1. Дослідження відношень на множині концептів «УСМ – УЖМ»

Під час перекладу з української жестової мови на українську словесну мову проблеми зняття омографії відрізняються від подібних проблем під час перекладу з однієї словесної мови на іншу. Деякі поняття, що однозначно сприймаються в словесній мові, у жестовій мові набувають кількох семантичних значень, які повинні бути уточнені для правильного перекладу. Це властиво не лише українській жестовій мові, а й деяким іншим жестовим мовам, зокрема, російській [91, 92]. Окрім того, часто трапляються випадки, коли один жест виражає ціле словосполучення.

Під час розроблення засобів машинного перекладу УЖМ постають додаткові проблеми, пов'язані з відсутністю необхідних словників української жестової мови та досконалого вивіреного опису граматики УЖМ.

Переклад з української жестової мови на українську словесну мову передбачає синтаксичний розбір речень жестовою мовою, побудову правил

перекладу української жестової мови на текст і навпаки, усунення омографії, уточнення змісту та синтез правильних конструкцій словесною мовою. На основі проведеного аналізу наявних сучасних систем перекладу жестових мов, автор розробив структуру системи перекладу української жестової мови на текст та навпаки, і визначено місце вирішених у роботі задач у загальній системі. Розроблена структура розбита на модулі, зв'язок між якими зображено структурною схемою (рис. 2.1).

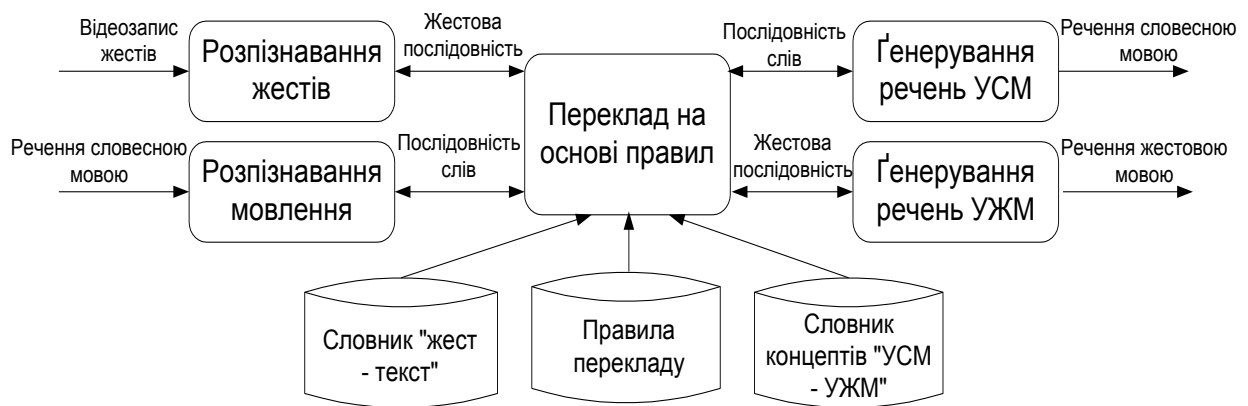


Рис. 2.1. Структурна схема системи перекладу української жестової мови

На вхід перекладу подається послідовність жестів у вигляді записаних глосів [31] (один глос відповідає одному жесту) або послідовність слів, залежно від того, з якої мови на яку здійснюється переклад. Для запису жестів у вигляді глосів використовуються слова, записані великими літерами. Наприклад, слово «бачити» українською жестовою мовою записано так: «БАЧИТИ». Модуль машинного перекладу на основі правил трансформує вхідну послідовність жестів у вихідну послідовність слів української мови. Цей модуль перекладає за допомогою спеціальних правил перекладу, словника «жест-текст» та словника концептів «УСМ – УЖМ». На виході процесу отримуємо речення українською словесною мовою. Система машинного перекладу також здійснює зворотній переклад, тобто з української словесної на українську жестову мову.

Для вирішення проблеми лексичної багатозначності слів укладено словник концептів «УСМ – УЖМ», для якого розроблено інфологічну модель. 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»).

Інфологічна модель розроблена на основі аналізу лінгвістичних відношень між концептами українською словесною та українською жестовою мовами. Адекватність інфологічної моделі досліджено з використанням корпусу паралельних речень «УСМ-УЖМ» [12, 13].

Для якіснішого перекладу української жестової мови потрібно сформулювати перелік концептів, тому доцільно ці концепти згрупувати та розробити на їхній основі онтологію українських словесної та жестової мов.

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. Оскільки онтологічні словники не описують повною мірою мовні конструкції для вираження семантичних зв'язків між поняттями, виникає необхідність розроблення деталізованіших моделей онтологій, зокрема граматично доповненої онтології. Для цього вона має містити більше інформації про граматичні зв'язки. Розроблення такої граматично доповненої онтології та її використання в системі машинного перекладу української жестової дає змогу частково вирішити проблему усунення багатозначності слів та підвищити оцінку якості перекладу.

Дослідження відомих методів вирішення проблеми багатозначності слів з використанням онтологій показало, що відомі методи обмежені лише контекстом слова і не надають додаткових переваг для синтаксичного і семантичного розбору речення. Для вирішення цієї проблеми розроблено математичну модель граматично доповненої онтології. Ця модель використана для семантично-синтаксичного розбору речень української мови.

Для практичного використання розробленої дисертантом математичної моделі граматично доповненої онтології необхідним було створення інформаційної технології, яка б давала змогу описувати цю онтологію, та перетворювала її на форму, придатну для використання у системах машинного перекладу.

Основним завданням під час перекладу української жестової мови, як і у всіх інших випадках перекладу з однієї мови на іншу, є правильне передавання змісту тексту, що перекладається. Для цього необхідною передумовою є повне розуміння вихідного тексту, що загалом є доволі складним завданням [93]. Для вирішення проблем, які виникають під час побудови систем машинного перекладу тексту на жестову мову та навпаки, важливо визначити відношення між концептами словесної та жестової мов, які передають зміст повідомлення [94]. Треба врахувати, що один жест жестової мови може виражати ціле словосполучення, та навпаки, деякі слова потребують пояснення у вигляді декількох жестів. Крім того, більшість висловлювань жестовою мовою набувають кілька значень, які треба уточнювати у словесній мові.

Ураховуючи особливості відношень між словами, жестами й концептами, виокремлено чотири основні ситуації, які потрібно враховувати під час перекладу з української жестової мови на українську словесну мову та навпаки.

1. Концепт подається одним словом УСМ та одним жестом УЖМ (табл. 2.1).

У цьому випадку слово безпосередньо перекладають одним жестом. Переклад простий, тому що одному жесту з УЖМ відповідає одне значення.

Таблиця 2.1

Одне слово відповідає одному жесту

№	Українська жестова мова	Українська словесна мова
1	Я ЛЮБИТИ ВІНОГРАД ЗЕЛЕНИЙ	Я люблю зелений виноград
2	Я ЖИТИ МІСТО	Я живу в місті

2. Концепт, який подається декількома словами, відповідає одному жесту (табл. 2.2).

Таблиця 2.2

Декілька слів відповідають одному жесту

№	Українська жестова мова	Українська словесна мова
1	Я НЕСТИ_РЮКЗАК	Я несу рюкзак
2	СЕРЦЕБИТТЯ	Серце б'ється рівно

Наприклад, висловлювання УСМ «серце б'ється рівно» перекладають за допомогою одного жесту. Приклад зображений на рис. 2.2.



Рис. 2.2. Зображення жесту «серце б'ється рівно»

3. Концепт, який подають одним словом, показують кількома жестами.

Така ситуація виникає в таких випадках:

- 1) переклад дієслів, коли концепт, який відповідає одному дієслову, генерує декілька жестів;
- 2) переклад загальних і специфічних іменників;
- 3) переклад іменника у множині;

4) переклад іменника з урахуванням статі.

Під час перекладу дієслів українською жестовою мовою треба враховувати, що в українській словесній мові існують три форми часу: минулий, теперішній та майбутній. Для того, щоб виразити дієслово теперішнього часу в українській жестовій мові використовують інфінітив, а для вираження дієслів минулого або майбутнього часів використовуються інфінітив та допоміжні жести «був» та «буду» відповідно (табл. 2.3). Наприклад, концепт УСМ «говорив» відповідає жестах «говорити + був», а концепт «говоритиму» – жестах «говорити + буду».

У жестовій мові трапляються ситуації, коли немає відповідного жесту для певного концепту, який можна виразити одним словом української словесної мови. Наприклад, в українській жестовій мові є відповідні жести для позначення слів «золото», «срібло», «мідь» тощо, але немає жесту для позначення слова «метал». Те саме стосується жесту для позначення слова «меблі»: є жести для слів «стіл», «стілець», «ліжка». Тому для вирішення цієї проблеми в жестовій мові використовують декілька жестів для позначення загального поняття (табл. 2.3).

Таблиця 2.3

Одне слово показують кількома жестами

№	Українська жестова мова	Українська словесна мова
1	Я КУПИТИ БУВ ХЛІБ	Я купила хліб
2	Я (НЕ МАЮ) СТІЛ СТІЛЕЦЬ ЛІЖКО БУДИНОК МІЙ	Я не маю меблів у своєму будинку
3	Я ЛЮБИТИ ПТАШКА ПТАШКА	Я люблю пташок
4	ВІН МАТИ (ЦУКЕРКА 2 РУКАМИ)	Він має цукерки
5	Я БАЧИТИ БАГАТО ДЕРЕВА ЗЕЛЕНИЙ	Я бачу зелені дерева
6	МОЯ БАБУСЯ МАТИ (КОЗА + ЖІНКА)	Моя бабуся має козу

Для перекладу жесту у множині є кілька способів: повторення жесту, використання додаткових жестів «багато», «різноманітні», «декілька» або показ жесту обома руками. Наприклад, слово «цукерки» можна показати такими способами: дворучним жестом «ЦУКЕРКИ», повторенням жесту «ЦУКЕРКА ЦУКЕРКА» чи «ЦУКЕРКА+БАГАТО». Кілька прикладів наведено на рис. 2.3 [95].



а



б

Рис. 2.3. Зображення жестів: а – «багато»; б – дворучного жесту «цукерки»

Для позначення статі об'єкта чи суб'єкта до основного жесту додається додатковий жест «чоловік» або «жінка». Здебільшого подібні уточнення статі робити немає необхідності, тому що це ускладнюватиме переклад. Проте у деяких випадках уточнення статі є необхідним для правильного передавання змісту (табл. 2.3).

4. Декілька концептів (декілька слів) відповідають декільком жестам (табл. 2.4).

Таблиця 2.4

Декілька слів відповідають декільком жестам

№	Українська жестова мова	Українська словесна мова
1	ДОРОГА РІВНО	Дорога йде рівно
2	ЧАС ЛЕТИТЬ	Час йде швидко

Згаданий випадок є найскладнішим із погляду перекладу, оскільки треба згенерувати жести УЖМ, які б відповідали за змістом декільком словам словесної мови. Наприклад, дієслово «йти» залежно від суб'єкта дії показують різними жестами: «Дорога йде рівно» і «Час йде швидко». У цьому разі жест-дієслово залежить від суб'єкта дії.

2.2. Інфологічна модель словника відношень між концептами та їхнім поданням в українській словесній та жестовій мовах

На основі аналізу можливих випадків відповідності жестових сполучень, словосполучень і концептів розроблено інфологічну модель словника відношень між концептами. Модель містить такі сутності: «Концепт», «Словосполучення», «Слово у словосполученні», «Граматична категорія слова», «Граматичне значення», «Граматична форма слова», «Початкова форма слова», «Жестове сполучення», «Жест у сполученні», «Граматична категорія жесту», «Граматична форма жесту», «Початкова форма жесту» та подана діаграмою «сутність-зв'язок» (рис. 2.4) [96].

Сутність «Концепт» у цій моделі відповідає певному висловлюванню (предмета, дії, ознаці тощо), яке має подання в українській словесній мові та в українській жестовій мові. Ця сутність має такі атрибути: «№ Концепту» (первинний ключ) та «Тлумачення» концепту (табл. 2.5).

Таблиця 2.5

Концепт		
№ концепту	№ словосполучення	Тлумачення
1	1	Прийшла осінь
2	2	Дівчина з'їла яблуко

Концепт в українській словесній мові виражається одним словом або словосполученням. Одне слово задається, як вироджене словосполучення, яке складається з одного слова. Словосполучення – це смислове і граматичне поєднання одного або декількох повнозначних слів.

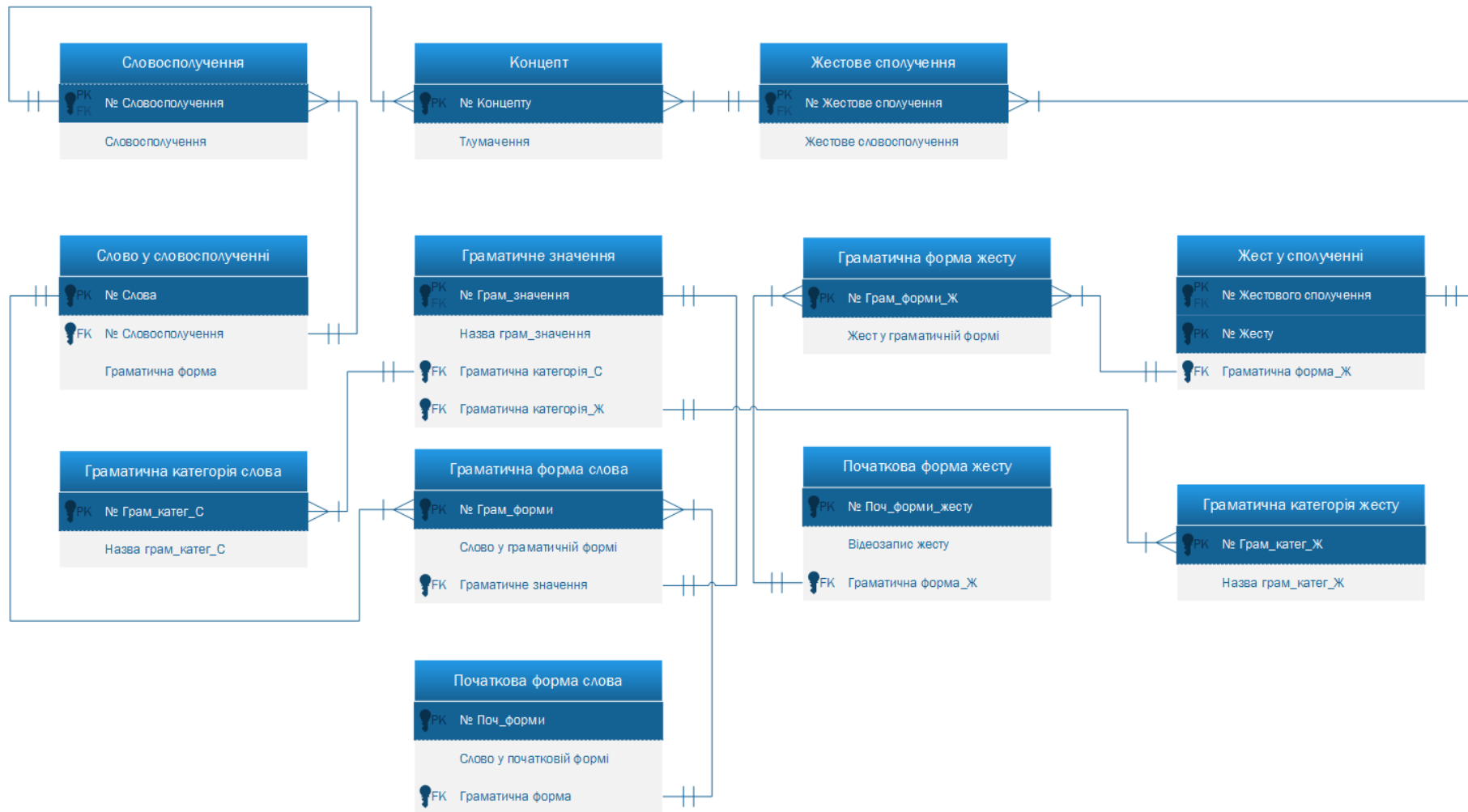


Рис. 2.4. Діаграма «сутність-зв'язок» словника концептів української словесної та жестової мов

Сутність «Словосполучення» має такі атрибути: «№ Словосполучення» (первинний ключ) та «Словосполучення». Наприклад, перелік значень для відношення Словосполучення (№ Словосполучення, Словосполучення) = {(1, Прийшла осінь), (2, Дівчина з'їла яблуко)} (табл. 2.6).

Таблиця 2.6

Словосполучення	
№ словосполучення	Словосполучення
1	Прийшла осінь
2	Дівчина з'їла яблуко

Сутність «Слово у словосполученні» містить атрибути «№ Слова» (порядковий номер слова у словосполученні), «№ Словосполучення», які є складеним первинним ключем (табл. 2.7).

Таблиця 2.7

Слово у словосполученні		
№ словосполучення	№ слова	№ Граматичної форми слова
1	1	4
1	2	5
2	1	6
2	2	7
2	3	2

Кожне слово у словосполученні вказане в певній граматичній формі. Граматична форма слова – це написання слова у конкретній формі плюс сукупність граматичних значень, які цьому написанню відповідають [97]. Наприклад, граматична форма слова «зеленого» відповідає граматичним значенням: родовий відмінок, чоловічий рід, число одинна. Сутність «Граматична форма слова» містить ключовий атрибут «№ Грам_форми» та атрибут «Слово у граматичній формі» (табл. 2.8).

Різні граматичні форми одного слова відповідають початковій формі слова. Початкова форма відповідає таким граматичним значенням: називному відмінку однини – для іменників та інфінітиву – для дієслів.

Таблиця 2.8

Граматична форма слова

№	Слово у граматичній формі	Граматичне значення
1	Хлопцями	5, 9, 10
2	Яблуко	1, 8, 12
3	Йдуть	9, 14, 17, 18, 23
4	Прийшла	8, 13, 16, 18, 22
5	Осінь	1, 8, 11
6	Дівчина	1, 8, 11
7	З'їла	8, 13, 16, 18, 23

Сутність «Початкова форма слова» містить ключовий атрибут «№ Поч_форми» та атрибут «Слово в початковій формі» (табл. 2.9).

Таблиця 2.9

Початкова форма слова

№	Слово у базовій формі	Граматичне форма слова
1	Хлопець	1
2	Йти	3
3	Прийти	4
4	Яблуко	2
5	Осінь	5
6	Дівчина	6
7	Їсти	7

Граматична форма слова може характеризуватися одним чи кількома граматичними значеннями. Граматичне значення – узагальнене (абстрактне) мовне значення, яке властиве групам слів, словоформ, синтаксичних конструкцій. Наприклад, різні за своїм значенням іменники «розум», «дуб», «хлопець» мають однакові граматичні значення: називного відмінка, числа однини, виражають належність до чоловічого роду. Сутність «Граматичне значення» містить атрибут «Назва грам_значення», що є первинним ключем (табл. 2.10).

Споріднені граматичні значення об'єднуються у граматичні категорії. Наприклад, чоловічий, жіночий і середній рід об'єднуються в одну категорію роду.

Таблиця 2.10

Граматичне значення		
№	Назва	Граматична категорія
1	Називний відмінок	1
2	Родовий відмінок	1
3	Давальний відмінок	1
4	Знахідний відмінок	1
5	Орудний відмінок	1
6	Місцевий відмінок	1
7	Кличний відмінок	1
8	Однина	2
9	Множина	2
10	Чоловічий рід	3
11	Жіночий рід	3
12	Середній рід	3
13	Минулий час	4
14	Теперішній час	4
15	Майбутній час	4
16	Доконаний вид	5
17	Недоконаний вид	5
18	Дійсний спосіб	6
19	Умовний спосіб	6
20	Наказовий спосіб	6
21	Перша особа	7
22	Друга особа	7
23	Третя особа	7

Деякі граматичні категорії мають і граматичне і семантичне значення, а деякі мають лише граматичне. Така семантична властивість може залежати й від самого слова. Наприклад, рід для живих істот ссавців має семантичне значення, а для неживих предметів – лише граматичне.

Граматичні категорії можуть бути спільними для декількох частин мови. Категорія числа охоплює багато частин мови (іменник, прикметник, займенник, дієслово), так само й категорія роду, а категорії особи, часу, виду, способу властиві тільки дієсловом, що пояснюється самим виявом предметів і явищ. Граматична категорія своїх показників не має, вона проявляється через граматичні значення, засобами вираження яких є граматичні форми. Сутність

«Граматична категорія слова» містить ключовий атрибут «№ Грам_катег_С» та атрибут «Назва грам_катег_С». Наприклад, перелік значень для відношення Граматична категорія слова (№ Грам_катег_С, Назва грам_катег_С) = {(1, Відмінок), (2, Число), (3, Рід), (4, Час), (5, Вид), (6, спосіб), (7, Особа)} (табл. 2.11).

Таблиця 2.11

Граматична категорія слова	
№	Назва
1	Відмінок
2	Число
3	Рід
4	Час
5	Вид
6	Спосіб
7	Особа

У жестовій мові концепт виражається одним або декількома синонімічними «Жестовими сполученнями». Жестове сполучення має два атрибути: «№ Жест_сполучення» (первинний ключ) та «Відеозапис» жестового сполучення (табл. 2.12).

Таблиця 2.12

Жестове сполучення	
№ жестового сполучення	Словосполучення
1	ПРИЙШЛА ОСІНЬ
2	ДІВЧИНА З'ЇЛА ЯБЛУКО

Жестове сполучення може складатися з одного або декількох жестів. Наприклад, перелік значень для відношення Жестове сполучення (№ Жест_сполучення, Відеозапис) = {(1, Відеозапис_1), (2, Відеозапис_2)}. Сутність «Жест у сполученні» містить ключовий атрибут «№ Жесту» та атрибут «Жест» (запис жесту у вигляді глоса) (табл. 2.13).

Граматична категорія жесту. В українській мові такі граматичні категорії, як рід, число, відмінок, час і особа передаються за допомогою закінчень. У

жестовій мові немає закінчень, які у словесній мові передають відтінки думок, час, особу тощо, проте граматичні категорії є. Це категорії числа, роду – для іменників та категорії часу, виду, способу – для дієслів.

Таблиця 2.13

Жест у сполученні

№ жестового сполучення	№ жесту	№ Граматичної форми жесту
1	1	4
1	2	5
2	1	6
2	2	7
2	3	2

Сутність «Граматична категорія жесту» містить ключовий атрибут «№ Грам_катег_Ж» та атрибут «Назва грам_катег_Ж» (табл. 2.14). Сутність «Граматична форма жесту» має наступні атрибути «№ Грам_форми_Ж» (первинний ключ) та «Жест у граматичній формі» (табл. 2.15).

Таблиця 2.14

Граматична категорія жесту

№	Назва
1	Число
2	Рід
3	Час
4	Вид
5	Спосіб

Таблиця 2.15

Граматична форма жесту

№	Жест у граматичній формі	Граматичне значення
1	ХЛОПЦІ	1, 9, 10
2	ЯБЛУКО	1, 8, 12
3	ЙДУТЬ	9, 14, 17, 18
4	ПРИЙШЛА	8, 13, 16, 18
5	ОСІНЬ	1, 8, 11
6	ДІВЧИНА	1, 8, 11
7	З'ЇЛА	8, 13, 16, 18

Сутність «Початкова форма жесту» має такі атрибути «№ Поч_форми_жесту» (первинний ключ) та «Відеозапис» показу жесту (табл. 2.16).

Таблиця 2.16

Початкова форма жесту		
№	Жест у базовій формі	Граматичне форма жесту
1	ХЛОПЕЦЬ	1
2	ЙТИ	3
3	ПРИЙТИ	4
4	ЯБЛУКО	2
5	ОСІНЬ	5
6	ДІВЧИНА	6
7	ЇСТИ	7

Отже, під час перекладу українською жестовою мовою враховується залежність граматичного значення жесту від граматичного значення слова.

Для тестування системи машинного перекладу використано граматичний парсер [98], який будує дерево граматичного розбору речення. Для перекладу української жестової мови застосовуються граматичні правила перекладу, описані у роботі [1, 2] та проводиться пошук для заміни кожного слова (словосполучення) українською словесною мовою на відповідний жест із використанням інфологічної моделі концептів.

Приклад дерева синтаксичного розбору речення «Хлопець несе рюкзак зі школи» наведено на рис. 2.5.

Розгляньмо як здійснюється переклад з використанням концептів для наведеного речення (табл. 2.17). Слово «хлопець» відповідає жесту «ХЛОПЕЦЬ», а словосполучення з двох слів «несе рюкзак» є концептом, який перекладається одним жестом «НЕСТИ_РЮКЗАК». Оскільки в жестовій мові прийменники не позначають (їх вживають лише в калькованій жестовій мові), то ми його не перекладаємо. Щоб розрізнити словосполучення «зі школи» і «до школи» в жестовій мові вказують напрямок руху. Слово «школа» має відповідник у жестовій мові «ШКОЛА». Зрозуміло, що якщо кожне

словосполучення «зі школи», «з магазину», «з бібліотеки» тощо, «до школи», «з магазину», «з бібліотеки» тощо записувати окремо, то ми отримаємо великий набір однакових за структурою словосполучень, які можна виокремити в групу «зі (до) школи (магазину, бібліотеки тощо)». Виокремлення таких груп можна використати для розроблення онтології з граматичною компонентою для української жестової та словесної мов.

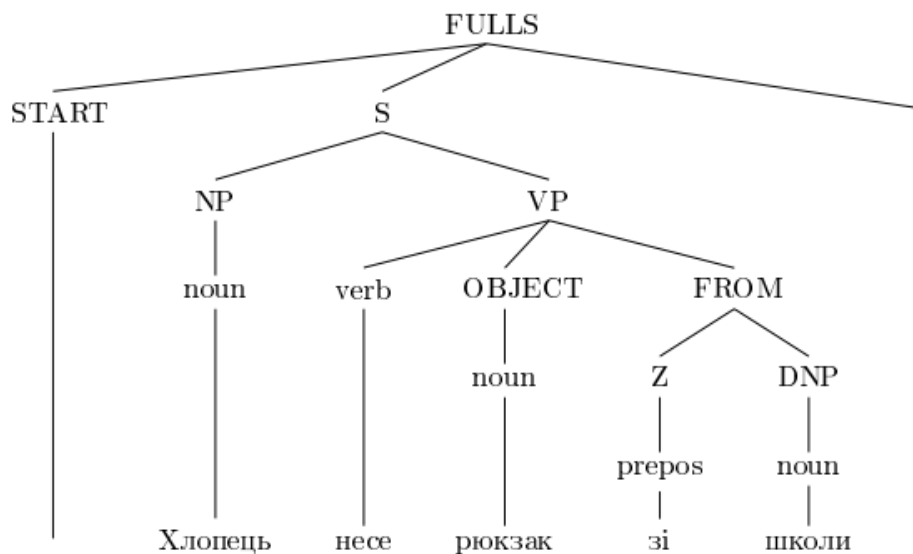


Рис. 2.5. Дерево граматичного розбору речення «Хлопець несе рюкзак зі школи»

Позначення: FULLS – речення; START – початок речення; S – основа речення;

VP – група дієслова; NP – підмет; DNP – додаток; . – кінець речення

Таблиця 2.17

Переклад із використанням концептів

№	Українська жестова мова	Українська словесна мова
1	ХЛОПЕЦЬ	Хлопець
6	НЕСТИ_РЮКЗАК	Несе рюкзак
10	ШКОЛА	Зі школи

Для тестування системи машинного перекладу УЖМ на основі правил із використанням словника концептів взято речення з корпусу паралельних текстів

«Українська словесна мова – Українська жестова мова», який містить 220 простих речень різної тематики українською словесною мовою та їх перекладів жестовою, 932 слова УСМ і 839 слова УЖМ. Основні характеристики корпусу наведено в табл. 2.18.

Таблиця 2.18

**Статистика корпусу паралельних текстів «Українська словесна мова –
Українська жестова мова»**

	УСМ	УЖМ
Кількість речень	220	
Кількість слів	932	839
Словник	432	398

Словник концептів розроблений згідно з описаними відношеннями між концептами та містить 50 жестових концептів та їх відповідних перекладів українською словесною мовою. Фрагмент словника концептів «УСМ–УЖМ» наведено в табл. 2.19.

Таблиця 2.19

Фрагмент словника концептів «УСМ–УЖМ»

№	Українська жестова мова	Українська словесна мова
1	КОЗА + ЖІНКА	Коза
2	СТІЛ + СТИЛЕЦЬ + ЛІЖКО	Меблі
3	БАГАТО + ЯБЛУКО	Яблука
4	ДОРОГА РІВНО	Дорога йде рівно
5	СЕРЦЕБИТТЯ	Серце б'ється швидко
6	НЕСТИ_РЮКЗАК	Несу рюкзак
7	ЧАС_ЛЕТИТЬ	Час летить
8	ЛІТАТИ	Літати

Для оцінювання якості системи перекладу з української словесної на українську жестову мову використано оцінки (word error rate) – оцінка помилок перекладу, PER (position-independent word error rate) – позиційно-незалежна оцінка помилок. Результати оцінювання якості перекладу без використання словника концептів та з його використанням наведені у табл. 2.20. Використання двомовного словника концептів «УСМ–УЖМ» та правил визначення, до якого концепту належить слово в певному контексті для української жестової мови, дало змогу покращити результати системи машинного перекладу на основі правил з 42,83 % до 34,58 % WER (що менша WER, то кращий результат).

Переклад покращився на 40 реченнях, які містили концепти. Наприклад, речення «Серце б'ється швидко» з використанням словника концептів перекладено на УЖМ «СЕРЦЕБИТТЯ» (швидкий темп жесту), а без використання словника концептів – «СЕРЦЕ БИТИ ШВИДКО» (у цьому випадку жест «БИТИ» означав «завдавати ударів»), тобто під час перекладу було повністю втрачено зміст речення.

Таблиця 2.20

Оцінка якості перекладу

	WER, %	PER, %
УСМ→УЖМ (без використання словника концептів)	42,83	18,35
УСМ→УЖМ (з використанням словника концептів)	34,58	14,65

Проведені експерименти підтверджують попередньо сформульовану гіпотезу про те, що для якіснішого перекладу української жестової мови належить укласти перелік концептів (в окремих випадках концепти становлять ціле речення). Для зменшення відсотку помилково перекладених слів у реченні необхідно погрупувати концепти та розробити на їх основі онтологію українських словесної та жестової мов.

2.3. Математична модель граматично доповненої онтології

З метою повнішого використання онтологічних словників у системах машинного перекладу необхідно розробити математичну модель для розроблення засобів вираження семантичних зв'язків у формі конструкцій речення.

Для виконання цього завдання необхідно, щоб онтологія була доповнена інформацією про граматичні зв'язки. Для використання у системах перекладу граматично доповнена онтологія повинна містити додаткові засоби для:

- 1) вираження граматичних атрибутів слів (граматичної категорії, граматичного значення);
- 2) опису основних граматичних конструкцій та їх зв'язку з поняттями.

Під час використання відомих онтологій для усунення багатозначності слів використовуються статистичні моделі, у яких враховується частота появи слів разом у певному контексті [99]. Такий підхід може забезпечити якісний переклад лише за умови використання великих навчальних корпусів. Використання онтологій для перекладу між мовами, для яких немає великих двомовних корпусів, потребує розроблення альтернативного підходу, який зорієнтований на граматичні зв'язки та їх семантичне значення. Автору не відомі опубліковані раніше наукові роботи, у яких моделі загальних онтологій були б доповнені такими граматичними зв'язками. Загалом є відомими літературі відомі декілька альтернативних моделей онтологій, таких як модель на основі семантичної мережі фреймів, машинна модель онтології та ін. [100].

За основу обрано математичну модель онтології, подану в [101], що визначається як кортеж

$$O = \langle C, M, R_c \rangle, \quad (1)$$

де $L = \{w_i\}$ $C = \{c_i\}$ – набір понять (концептів) предметної галузі;

$M_i = \{m_{1i}, \dots, m_{di}\}$ – множина атрибутів поняття c_i (d – кількість атрибутів, що описують це поняття);

R_c – множина відношень на множині понять (гіпонімії, гіперонімії, меронімії, голонімії тощо).

Означимо пропоновану математичну модель граматично доповненої онтології як кортеж

$$O_G = \langle O, P, S, T, R_p \rangle, \quad (2)$$

де $P = \{p_i\}$ – множина предикатів;

$S = \{s_i\}$ – множина висловлювань, які можуть бути задані переліком або граматикою, де $s_i = ((w_1, g_1), (w_2, g_2), \dots, (w_n, g_n))$ – рядок граматично доповнених термінів, $n = \text{Len}(s_i)$ – довжина рядка, (w_i, g_i) – граматично доповнений термін;

$T = \{t_i\}$ – множина параметризованих висловлювань, де $t_i = (s_k, f_i)$ – пара висловлювання-функція параметризації;

$f_i : \{1, 2, \dots, \text{Len}(s_i)\} \rightarrow \{0, 1, \dots, r\}$ – функція параметризації, r – кількість параметрів параметризованого висловлювання t_i , $f(k) > 0$ означає, що відповідний термін у позиції k є параметром з номером $f(k)$, і може бути замінений на інший термін із множини гіпонімів;

$I \subset T \times P$ – відношення інтерпретації між параметризованими висловлюваннями і предикатами;

R_p – множина відношень на множині предикатів (гіпонімії, гіперонімії, меронімії тощо).

Таке означення граматично доповненої онтології дає змогу виразити зв'язок між поняттями, предикатами і засобами їх вираження у вигляді мовних конструкцій. Наприклад, розглянемо предикат «вчити (кого, чому)». Застосуємо його у реченні «Я вчу школярів математиці». Це речення можна побудувати і таким чином «Я вчу математиці школярів». Таке речення не відповідає параметризованому висловлюванню згідно з наведеною математичною моделлю граматично доповненої онтології, тому що ми можемо поміняти параметри місцями (у наведеному прикладі, параметри «кого, чому»). Під параметризованим розуміємо таке висловлювання, у якому параметри

пронумеровані і відповідають номерам предметних змінних предиката. Наприклад, предиката «вчити (суб'єкт, кого, чому)» відповідає параметризоване висловлювання «1.хто вчити 2.кого 3.чому». Цьому параметризованому висловлюванню, як приклад, відповідає речення «Я вчу школярів математиці». Враховуючи граматично доповнені терміни у онтології, запишемо це параметризоване висловлювання з використанням відмінків (позначено С, casus) для параметрів предиката «1.С1 вчити 2.С4 3.С3». Цифри зліва від літери С позначають позицію параметра у параметризованому висловлюванні, а цифри справа позначають номер відмінка в українській мові (називний – 1, родовий – 2, давальний – 3, знахідний – 4, орудний – 5, місцевий – 6, кличний – 7). У реченні деякі параметри можуть опускатися. Наприклад, у реченні «Я вчу школярів» відсутній третій параметр «чому» висловлювання, який відповідає предиката «вчити (суб'єкт, кого, чому)».

Крім того, предикат «вчити» вживається й у іншому значенні «вчити що» (вивчати). У цьому випадку, згідно з наведеною математичною моделлю граматично доповненої онтології запишемо його так: «вчити (суб'єкт, що)», якому відповідає параметризоване висловлювання «1.С1 вчити 2.С4». Наприклад, речення «Я вчу вірш» відповідає цьому предикату.

Серед відношень на множині концептів, розглянемо відношення гіпонімії/гіперонімії. Наприклад, у реченнях «Я вчу школярів математиці» та «Я вчу людей математиці» між поняттями «школяр» та «людина» встановлено відношення гіпонімії/гіперонімії (поняття «школяр» є гопонімом поняттю «людина»). Дерево семантично-синтаксичного розбору для речення «Я вчу школярів математиці», коли розбір відповідає онтології (вага правила більша за 1), зображено на рис. 2.6. Під семантично-синтаксичним розбором розуміємо семантичний розбір, коли розбір здійснюється з використанням ГДО, і синтаксичний розбір, коли ГДО не використовується.

Отже, за допомогою відношень між поняттями у граматично доповненій онтології на місця параметризованих змінних можна поставити лише правильні по змісту поняття.

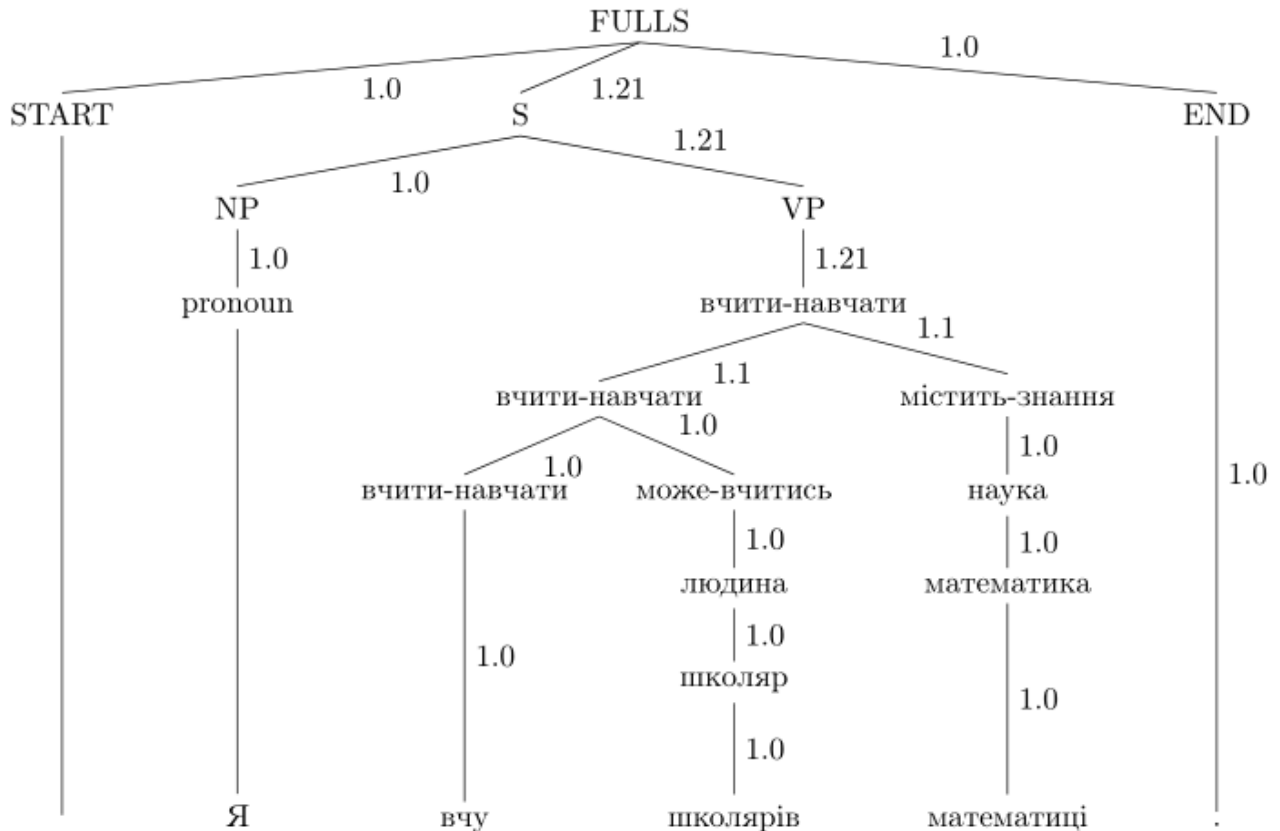


Рис. 2.6. Дерево семантично-синтаксичного розбору речення «Я вчу школярів математиці»

Так, речення «Я вчу дошку математиці» згідно з наведеною граматично доповненою онтологією є некоректним і для цього речення дерево семантично-синтаксичного розбору відповідає звичайному синтаксису речення (вага правила дорівнює 1) та зображено на рис. 2.7.

Відношення меронімії/голонімії у граматично доповненій онтології відображає відношення «частина-ціле». Це відношення властиве іменникам. Наприклад, двері є частиною (меронімом) будинку. Отже, у предикаті «1.C1 заходити у 2.C4» як параметр на 2 позиції можна підставити термін «двері» або термін «будинок».

Застосування розробленої математичної моделі у парсері української мови на основі ймовірнісних афіксних контекстно-вільних граматик дало змогу отримати семантичний розбір речення за умови відповідності речення правилам онтологічного словника.

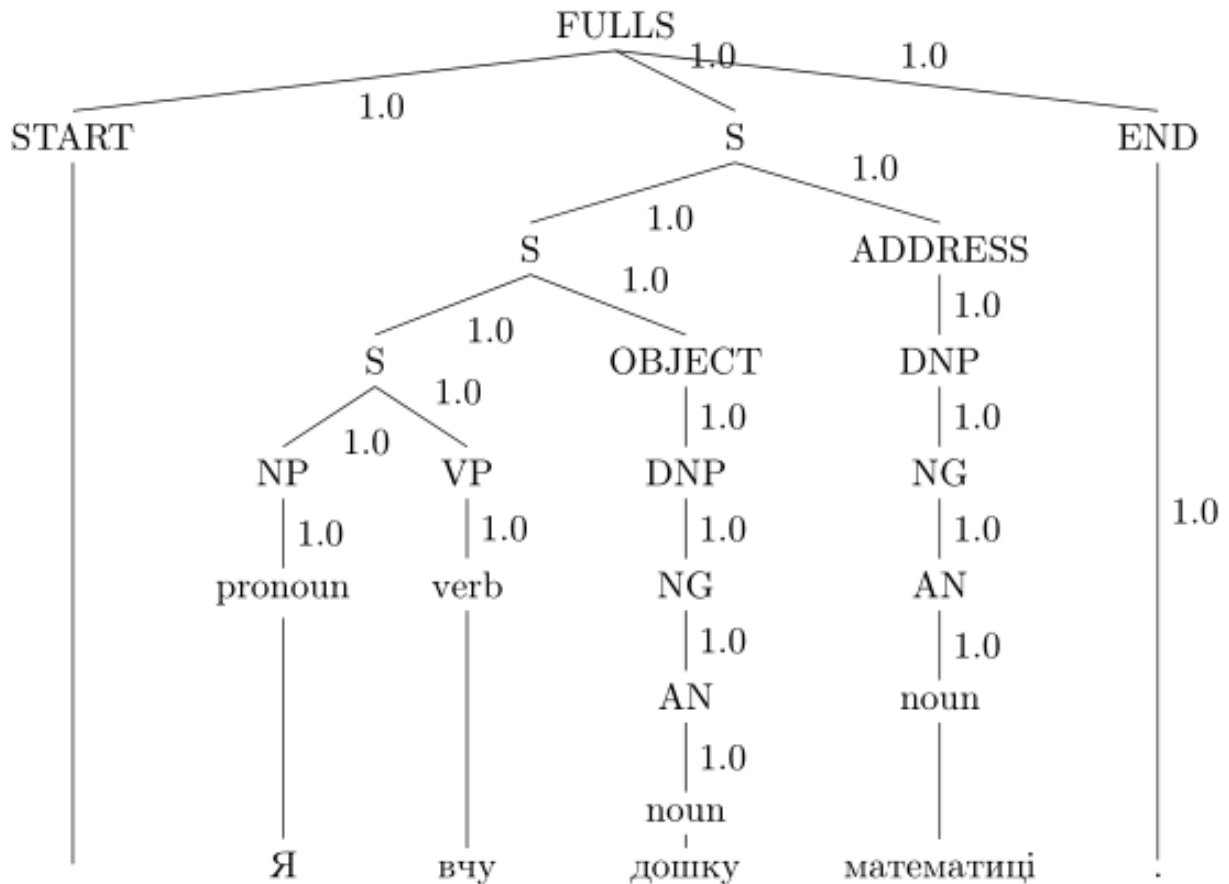


Рис. 2.7. Дерево семантично-синтаксичного розбору для речень «Я вчу дошку математиці»

2.4. Предметно-орієнтована мова опису граматично доповненої онтології

Важливим завданням є інтегрування граматично доповненої онтології в системи машинного перекладу. Для цього вона повинна містити більше інформації про граматичні зв'язки. Для ефективного усунення багатозначності слів під час машинного перекладу використано граматично доповнену онтологію, яка містить граматичні атрибути та мовні конструкції.

Практичне використання словників граматично доповнених онтологій у системах автоматизованого перекладу потребує розроблення засобів уніфікованого опису таких словників. Ці засоби мають надавати зручний спосіб редагування онтологій та можливість їх перевіряння на відповідність вимогам цілісності й несуперечливості.

Автором для опису словників граматично доповнених онтологій використано підхід на основі предметно-орієнтованих мов. Такий підхід дає змогу уніфікувати подання онтології у вигляді файлів опису із чітко визначеною структурою.

Розроблення предметно-орієнтованої мови опису онтологічного словника вимагає виконання таких завдань:

- 1) вибір засобів для створення предметно-орієнтованої мови;
- 2) розроблення синтаксису мови;
- 3) розроблення оболонки для редагування файлів опису онтологічного словника із перевіркою граматики і наданням користувачу контекстно-залежної підказки;
- 4) вибір засобів для перевірки (компіляції) файлів і перетворення їх на форму, придатну для швидкого пошуку.

Вибір системи побудови онтології залежить від практичного завдання (цілей розробника), галузі знань, у межах якої будується онтологія.

Створення онтології складається з таких етапів:

- 1) виділення понять;
- 2) формулювання відношень між поняттями;
- 3) наповнення онтології конкретними екземплярами;
- 4) формулювання аксіом (правил).

Відомі системи побудови онтологій DOE (Differential Ontology Editor) [102], Ontolingua [103], OntoEdit [104], WebOnto [105], Protégé [106] не мають засобів зв'язування суті понять і можливих граматичних конструкцій для їх вираження. Для можливості введення в онтологію граматичних конструкцій української словесної та жестової мов за активної участі автора розроблено предметно-орієнтована мова опису граматично доповненої онтології (Domain Specific Language, DSL). Термін предметно-орієнтована мова означає мову

програмування або моделювання, що застосовується для розв'язання конкретних задач в термінах, максимально наближених до даної предметної галузі [107].

Однією з проблем під час створення та подальшого використання DSL є наявність програмних засобів створення та підтримання предметно-орієнтованих мов (мовних інструментальних засобів або мовних інструментаріїв). Мовні інструментальні засоби – це спеціалізовані інтегровані середовища розроблення (integrated development environment, IDE) для визначення та створення DSL [108].

Практично всі сучасні середовища розроблення (наприклад, Eclipse IDE, Microsoft Visual Studio та ін.) мають гнучку архітектуру і дають змогу додавати підтримку нових предметно-орієнтованих мов.

Як приклад повноцінного мовного інструментарію можна розглядати програмний продукт JetBrains MPS [109]. Він призначений для створення DSL, а для розроблення програм на DSL в ньому використовується проєкційний редактор, який подає сценарій на DSL як синтаксичне дерево.

Існують застосунки для створення інтелектуальних редакторів, наприклад, IntelliJ IDEA Language API [110], Xtext [111] (на базі платформи Eclipse) та інші.

Застосунок Xtext забезпечує створення набору засобів для наповнення, редагування та верифікації DSL-мови, що містить аналізатор коду мови, інструмент форматування коду, компілятор, редактор коду та ін.

2.4.1. Формальне подання мови опису граматично доповненої онтології

Основою для побудови мови опису граматично доповненої онтології стала математична модель граматично доповненої онтології.

Наповнення граматично доповненої онтології відбувається за допомогою розробленої предметно-орієнтованої мови. Предметно-орієнтована мова UkrNet реалізована з використанням додатка Xtext середовища розроблення

програмного забезпечення Eclipse, який дає змогу для граматики створити мову DSL [112].

Розроблення предметно-орієнтованої мови UkrNet здійснено з врахуванням граматики цієї мови в Xtext. Потім, використовуючи засоби Xtext, створено синтаксичний аналізатор і текстовий редактор для мови UkrNet. Вигляд вікна текстового редактора для мови UkrNet зображено на рис. 2.8.

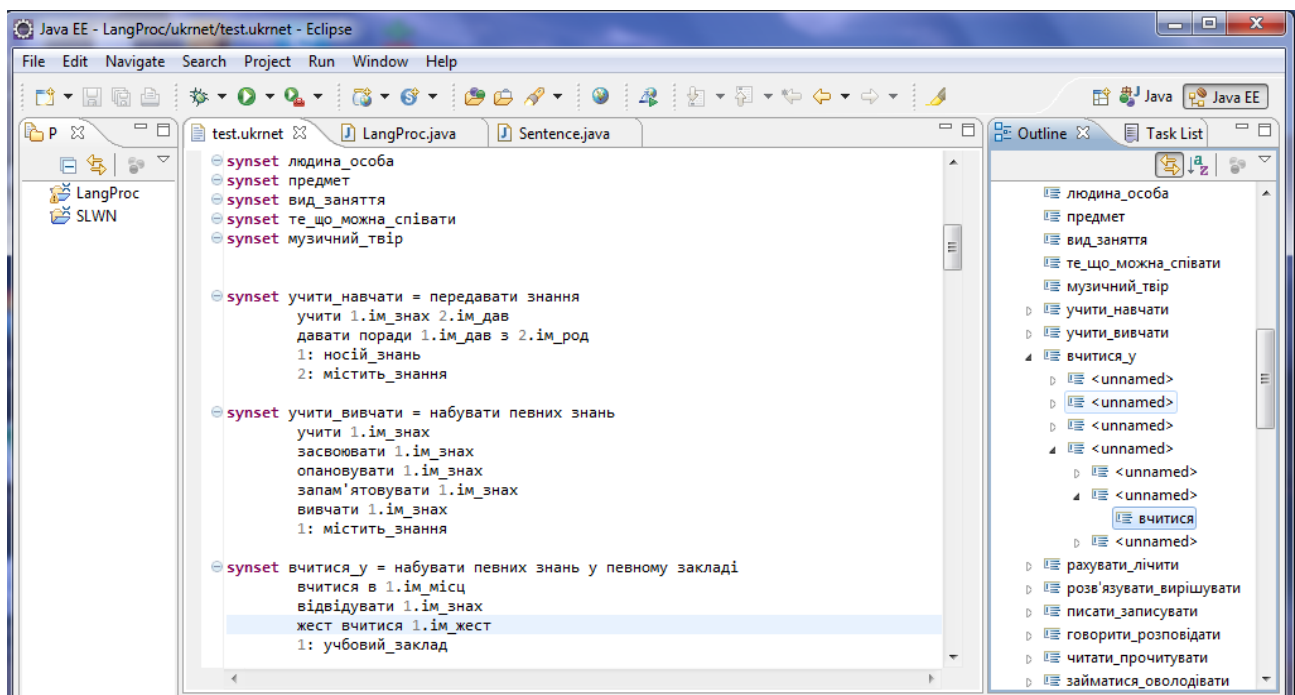


Рис. 2.8. Мова для опису граматично доповненої онтології UkrNet

Розглянемо опис понять граматично доповненої онтології.

Для позначення граматичних значень введено поняття *grammar_attribute* **<назва категорії> [= опис]**. Граматичні категорії можуть бути спільними для декількох частин мови. Категорія числа охоплює багато частин мови (іменник, прикметник, займенник, дієслово), так само й категорія роду, а категорії особи, часу, виду, способу властиві тільки дієсловам. Граматична категорія своїх показників не має, вона виявляється через граматичні значення, засобами вираження яких є граматичні форми. Для опису онтології з граматичною компонентою введено такі граматичні значення:

grammar_attribute ім_наз // іменник у називному відмінку

grammar_attribute ім_род // іменник у родовому відмінку
grammar_attribute ім_дав // іменник у давальному відмінку
grammar_attribute ім_знах // іменник у знахідному відмінку
grammar_attribute ім_оруд // іменник у орудному відмінку
grammar_attribute ім_місц // іменник у місцевому відмінку
grammar_attribute д_інф // дієслово (інфінітив)
grammar_attribute д_жест // дієслово (жест)
grammar_attribute ім_жест // іменник (жест)

Для позначення синсетів використовують такі описи:

synset <назва синсету> [= опис] // синсет без відповідних словесних реалізацій (просто поняття);

synset <назва синсету> [= опис]

учити 1.ім_знах 2.ім_дав // певне словесне вираження, напр. учити учнів плаванню

давати поради 1.ім_дав з 2.ім_род // інша словесна реалізація, напр. давати поради учням з плавання

1: носій_знань // вимоги до слова 1 (optional)

2: містить_знання // вимоги до слова 2 (optional)

Кожен синсет становить деяке значення, поняття 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 представляє один семантичний концепт.

Для систематизації іменників введено загальне поняття до синсету1: **common**(<назва синсету1>) = <назва синсету2>, що дає змогу побудувати родо-видове відношення, відоме як гіперонімія, та обернене йому відношення гіпонімії. Наприклад, синсет «книга» є гіпонімом синсету «носій_знань»:

common(книга) = носій_знань

Введення поняття **parts**(<назва синсету1>) визначає відношення «частина-ціле», або меронімію. Опис цього поняття має такий вигляд:

parts(<назва синсету1>) = <назва синсету2>, ..., <назва синсетуN> // синсет2,..., синсетN є невід’ємною частиною синсету1;

Наприклад, синсет «книга» завжди містить синсет «сторінка»:

parts(книга) = сторінка

Поняття **optional_parts**(<назва синсету1>) має такий опис:

optional_parts(<назва синсету1>) = <назва синсету2>, ..., <назва синсетуN> // синсет2,..., синсетN є можливою частиною синсету1 (за звичного розуміння);

Наведене поняття подібне до відношення **parts**(<назва синсету1>) та використовується в тому випадку, коли *синсет2, ..., синсетN* є можливою частиною *синсету1*. Наприклад, синсет «сторінка» не завжди має синсет «номер_сторінки»:

optional_parts(сторінка) = номер_сторінки

Для позначення відношення однозначної асоціації між синсетами використовується поняття **associations**(<назва синсету1>), що має такий опис:

associations(<назва синсету1>) = <назва синсету2>, ..., <назва синсетуN> // синсет2,..., синсетN завжди асоціюються з синсет1;

Для прикладу, синсет «школяр» завжди асоціюється з синсетом «школа»:

associations(школяр) = школа // школяр завжди ходить у школу

Для позначення відношення можливої асоціації між синсетами введено поняття **optional_associations**(<назва синсету1>). Це поняття має такий опис в онтології:

optional_associations(<назва синсету1>) = <назва синсету2>, ..., <назва синсетуN> // синсет2,..., синсетN можуть асоціюватися з синсет1;

Це відношення використовується тоді, коли не завжди синсет1 асоціюється з синсетом2,..., синсетомN. Наприклад, синсет «заліковка» не завжди є у студента:

optional_associations(студент) = заліковка // часто студенти мають заліковки

Крім того, поняття може входити у різні класифікації:

common(учень) = людина // учень має всі ознаки людини

common(людина) = людина_біологічний_вид // людина має всі ознаки людина_біологічний_вид

Отже, учень має всі ознаки людина_біологічний_вид.

Зрозуміло, що учень не є більш конкретним представником людина_біологічний_вид:

parts(учень) = людина_біологічний_вид

Приклад неправильного відношення – школяр не є частиною учня, а є більш конкретним об'єктом.

optional_parts(учень) = школяр // неправильно!!!

common(школяр) = учень // правильно

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 [113] та словника жестів [114]. До кожного слова також зібрано всі його синоніми. Після цього в редакторі онтологій занесено ці слова у вигляді синсетів (синонімічних рядів) та описано граматичні конструкції, в яких використовується це слово. У синсет також включено граматичну конструкцію жестової мови. Так, ми зможемо отримати однозначний переклад певного словосполучення.

Наприклад, розглянемо слово «учити». У словосполученні слово «учити» може означати «навчати когось» (учити учнів математиці) або може означати «вивчати самому» (учити математику). Тому для цього слова у граматичній онтології записано два синсети:

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

1: носій_знань

2: містить_знання

synset **учити_вивчати** = набувати певних знань

учити 1.ім_знах

засвоювати 1.ім_знах

опановувати 1.ім_знах

запам'ятовувати 1.ім_знах

вивчати 1.ім_знах

жест вивчати 1.ім_жест 2.ім_жест

1: містить_знання

Також записано синсет «учитись у», що означає учитись у певному навчальному закладі:

synset **вчитися_у** = набувати певних знань у певному закладі

вчитися в 1.ім_місц

відвідувати 1.ім_знах

жест вчитися 1.ім_жест

1: навчальний_заклад

Висновки до другого розділу

У другому розділі розроблено інфологічну модель концептів для побудови системи перекладу української жестової мови на основі аналізу лінгвістичних відношень між концептами УСМ та УЖМ. Тестування цієї моделі проведено на прикладі корпусу паралельних речень «УСМ–УЖМ». Проведене оцінювання якості системи перекладу з використанням інфологічної моделі концептів показало покращення результату перекладу на тих реченнях, які містили концепти, з 42,83 % до 34,58 % WER. Недоліком системи є невеликий обсяг вхідних даних для тестування (220 речень).

Під час дослідження для перекладу української жестової мови згенеровано граматично правильні конструкції словесної мови, що відтворюють зміст жестового висловлювання. На основі проведених експериментів встановлено, що для якісного перекладу української жестової мови потрібно сформулювати перелік

концептів, тому доцільно ці концепти згрупувати та створити на їхній основі онтологію українських словесної та жестової мов.

Розроблено математичну модель граматично доповненої онтології, яка забезпечує можливість введення в онтологію граматичних категорій та граматичних значень, які виражені граматичними формами.

Застосування розробленої математичної моделі у парсері української мови на основі ймовірнісних афікських контекстно-вільних граматик дало змогу отримати семантичний розбір речення за умови відповідності речення правилам онтологічного словника.

Для забезпечення можливості введення в онтологію граматичних категорій розроблено спеціалізовану предметно-орієнтовану мову опису UkrNet граматично доповненої онтології. Предметно-орієнтована мова UkrNet реалізована з використанням додатка Xtext середовища розроблення програмного забезпечення Eclipse, який дає змогу для граматики створити предметно-орієнтовану мову. Граматично доповнена онтологія містить 1200 слів з різних предметних галузей. Ці слова занесені у вигляді синсетів (синонімічних рядів) та описані граматичні конструкції, в яких використовуються ці слова. У синсет також включено граматичні конструкції жестової мови. Це забезпечує отримання однозначного перекладу певного словосполучення. Загалом підхід, що ґрунтується на граматично доповненій онтології дає змогу покращити результат перекладу української жестової та словесної мов і може ефективно використовуватись у системах машинного перекладу.

РОЗДІЛ 3

МЕТОДИ ДВОСТОРОННЬОГО ПЕРЕКЛАДУ АНОТОВАНОЇ УКРАЇНСЬКОЇ ЖЕСТОВОЇ МОВИ

У цьому розділі проведено аналіз практичної застосовності таких методів машинного перекладу анотованої української жестової мови як: статистичного методу, методу на основі правил та методу з використанням ГДО. Описано основні правила перекладу анотованої української жестової мови на словесну та навпаки. Для семантично-синтаксичного розбору речень української словесної та анотованої української жестової мов використано метод розбору на основі афіксної ймовірнісної контекстно-вільної граматики (Affix Probabilistic Context-Free Grammar, APCFG), що, своєю чергою, дало змогу збільшити відсоток правильно синтаксично проаналізованих речень з 72 % до 91 % для речень УСМ та з 65 % до 90 % для речень УСМ. Для перекладу речень української словесної мови на речення анотованої української жестової мови використано метод трансформації вузлів дерева семантично-синтаксичного розбору, що дало змогу підвищити відсоток правильно перекладених речень з 57,2 % до 93,2 % (порівняно зі статистичним методом машинного перекладу).

3.1. Застосування статистичних методів машинного перекладу анотованої української жестової мови

Розроблення засобів машинного перекладу української жестової мови є складною проблемою, вирішення якої вимагає розв'язання низки задач, пов'язаних із розпізнаванням жестів та їх інтерпретацією, усуненням неоднозначності, граматичним розбором жестових та словесних речень, збереженням та використанням знань експертів-перекладачів.

Незважаючи на збільшення кількості наукових робіт, пов'язаних із синтаксичними особливостями української жестової мови [97, 115, 116], досі

немає відкритих для використання програми автоматичного синтаксичного аналізу речень українською жестовою мовою.

Статистичні системи машинного перекладу жестових мов використовують двомовні корпуси, які містять повні речення. Такі корпуси використовуються для навчання цих статистичних систем. Але коли йдеться про ЖМ, то виникають дві основні проблеми. Перша проблема – це відсутність великих за обсягом корпусів. Наявні корпуси використовують позначення глосів (один глос – один жест), які є занадто складними для вивчення [49]. Для української жестової мови розроблено корпус більш ніж на 220 речень, який містить речення на українській словесній мові та їх переклад на українську жестову мову, враховуючи основні правила перекладу. Друга проблема – це відсутність стандарту позначення жестів.

Для статистичного машинного перекладу української словесної мови на анотовану українську жестову мову використано модель IBM № 1 (розробила корпорація International Business Machines, IBM) та EM-алгоритм (Expectation-maximization) для вирівнювання слів у реченнях корпусу паралельних текстів «українська словесна мова – українська жестова мова».

Наприклад, для слова УЖМ «Ти (Ви)» існує кілька перекладів в УСМ, як «Ти (Ви)» або «Тебе (Вас)» або «Тобі (Вам)». Більшість слів мають кілька перекладів і деякі з них більш ймовірні, ніж інші. Тому для перекладу використовують статистичні дані на основі кількості слів в двомовному корпусі. Табл. 3.1 показує можливий переклад слова «Ти (Ви)».

Таблиця 3.1

Підрахунок кількості перекладів слова «Ти (Ви)»

Переклад «Ти (Ви)»	Кількість
«Ти (Ви)»	34
«Тебе (Вас)»	14
«Тобі (Вам)»	12

Це слово трапляється 60 разів в нашому двомовному корпусі та перекладається 34 рази у слово «Ти (Ви)», 14 разів – у слово «Тебе (Вас)» і 12 разів – у слово «Тобі (Вам)». Для оцінки ймовірностей перекладу скористаємось цими підрахунками і отримаємо максимальну вірогідність[117]:

$$p_f(e) = \left\{ \begin{array}{ll} 34/60 = 0.57 & \text{if } e = \text{«Ти(Ви)»}, \\ 14/60 = 0.23 & \text{if } e = \text{«Тебе(Вас)»}, \\ 12/60 = 0.2 & \text{if } e = \text{«Тобі(Вам)»} \end{array} \right\} \quad (1)$$

У зв'язку з відмінністю порядку слів у реченнях УСМ та УЖМ, потрібно, щоб слова були перерозподілені (вирівняні) під час перекладу. Наведемо приклад речення, у якому слова УСМ та УЖМ знаходяться в однаковому порядку:

УСМ	¹	²	³
	Я	читаю	книгу.
УЖМ	₁	₂	₃
	Я	ЧИТАТИ	КНИГА.

Усі моделі статистичного машинного перекладу засновані на принципі вирівнювання слів. Для вирівнювання позицій слів у реченні використовується функція вирівнювання, яка відображає кожне слово УСМ у позиції i у слово УЖМ у позиції j : $a: j \rightarrow i$ [117]. Для наведеного вище прикладу функція вирівнювання $a: \{1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3\}$.

УСМ	¹	²	³	⁴
	Я	читаю	цікаву	книгу.
				
УЖМ	₁	₂	₃	₄
	Я	ЧИТАТИ	КНИГА	ЦІКАВИЙ.

У наведеному прикладі бачимо, що не завжди порядок слів у реченнях УСМ та УЖМ однаковий. Для цього прикладу функція вирівнювання $a: \{1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 4, 4 \rightarrow 3\}$. Також під час перекладу для вирівнювання слів

виникають випадки, коли одному слову в УСМ може відповідати багато слів в УЖМ та навпаки (функція вирівнювання $a: \{1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 2, 4 \rightarrow 3\}$):

УСМ	1	2	3	
	Я	читав	книгу.	
		\		
УЖМ	1	2	3	4
	Я	ЧИТАТИ	БУВ	КНИГА.

Для перекладу питальних речень в українській жестовій мові використовується таке правило: питальні слова (наприклад, «де», «коли», «чому», «скільки»), завжди ставляться в кінці речення. Наприклад, питальне речення «Де ти живеш?» в УЖМ передаємо так (функція вирівнювання $a: \{1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 1, 4 \rightarrow 4\}$):

УСМ	1	2	3	4	
	Де	ти	живеш	?	
		/	/		
		\	\		
УЖМ	1	2	3	4	
	ТИ	ЖИТИ	ДЕ	?	

Найпростішою статистичною моделлю перекладу є модель дослівного перекладу, відома як модель IBM №1. Ця модель передбачає, що кожному слову у реченні УЖМ відповідає одне або нуль слів у реченні УСМ.

Модель IBM №1 визначає ймовірність перекладу речення УЖМ $f = (f_1, \dots, f_{l_f})$ довжиною l_f у речення УСМ $e = (e_1, \dots, e_{l_e})$ довжиною l_e з вирівнюванням кожного слова УСМ e_j у слово УЖМ f_i відповідно до функції вирівнювання $a: j \rightarrow i$ так [118]:

$$P(e, a | f) = \frac{\epsilon}{(l_f + 1)^{l_e}} \prod_{j=1}^{l_e} t(e_j | f_{a(j)}) \quad (2)$$

Якщо є вирівняний вручну корпус паралельних текстів, тоді можливо оцінити параметри моделі IBM №1 методом максимальної правдоподібності. Оскільки такого корпусу для української жестової мови не існує, тому застосовуємо EM-алгоритм [117]. Розглянемо застосування EM-алгоритму на

простому прикладі (рис. 3.1). У табл. 3.2 наведено кілька ітерацій на трьох реченнях корпусу з чотирма вхідними словами (Я, ЧИТАТИ, ЦІКАВИЙ, КНИГА) і чотирма вихідними словами (я, читаю, цікава, книгу). Спочатку ймовірності перекладу зі слова жестової мови на словесну $\frac{1}{4} = 0,25$. Враховуючи ці початкові ймовірності, обчислюємо ймовірності алгоритму EM.

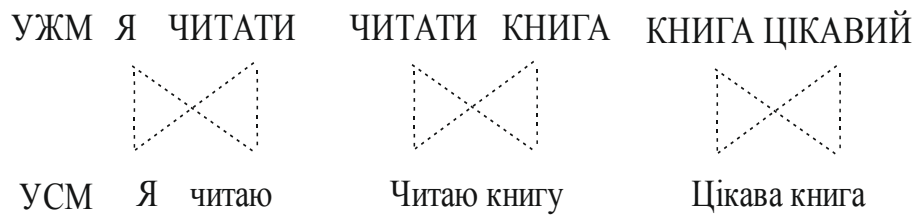


Рис. 3.1. Приклад речень для перекладу

Таблиця 3.2

Результати застосування EM-алгоритму

f	e	Ініціалізація	1-ша ітер.	2-га ітер.	3-тя ітер.	...	Кінець
Я	я	0.25	0.5	0.64	0.75	...	1
ЧИТАТИ	я	0.25	0.25	0.18	0.12	...	0
КНИГА	я	0.25	0.25	0.18	0.13	...	0
Я	читаю	0.25	0.25	0.18	0.12	...	0
ЧИТАТИ	читаю	0.25	0.5	0.64	0.75	...	1
ЦІКАВИЙ	читаю	0.25	0.25	0.18	0.13	...	0
ЧИТАТИ	цікава	0.25	0.5	0.43	0.35	...	0
ЦІКАВИЙ	цікава	0.25	0.5	0.57	0.65	...	1
Я	книга	0.25	0.5	0.43	0.35	...	0
КНИГА	книга	0.25	0.5	0.57	0.65	...	1

Візуально результати вирівнювання слів можна подати у вигляді матриці вирівнювання (рис. 3.2):

	Я	ЧИТАТИ	БУВ	КНИГА
я				
читав				
книгу				

	ТИ	ЖИТИ	ДЕ	?
де				
ти				
живеш				
?				

Рис. 3.2. Матриця вирівнювання слів

Слова речення УЖМ (рядки) вирівняні у слова речення УСМ (стовпчики), як показано у матриці вирівнювання. Вирівнювання не завжди може бути один-до-одного. На рис. 3.2 наведено приклад, коли одному слову в УСМ (читав) відповідає два слова в УЖМ (ЧИТАТИ БУВ). Блок-схема алгоритму статистичного перекладу речень української словесної мови на анотовану українську жестову мову зображено на рис. 3.3.

У результаті проведених досліджень встановлено, що до основних проблем комп'ютерного перекладу з української словесної мови на анотовану українську жестову мову та навпаки належать:

- неоднозначність перекладу, зумовлена тим, що кількість слів жестової мови відрізняється від кількості слів словесної (відео-словник УЖМ містить близько 2000 жестів, що зображають слова різної тематики [114], а відомий тлумачний «Словник української мови» – 135000 слів [113]);
- граматику жестової мови відрізняється від граматики словесної мови;
- існує значна відмінність у засобах передавання змісту речення (у жестовій мові велике значення має посліовність слів у реченні);
- виникає необхідність використання в УМЖ додаткових жестів для уточнення змісту речення;

- необхідність використання поряд із жестами дактильної абетки вказівних жестів, транслітерації власних назв та термінів.

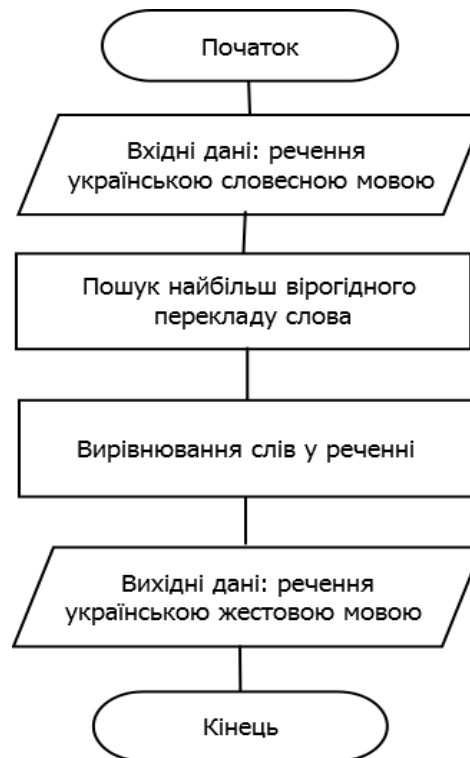


Рис. 3.3. Блок-схема алгоритму статистичного перекладу речень УСМ у речення анотованої УЖМ

3.2. Застосування методів машинного перекладу анотованої УЖМ на основі правил

Структурну основу двоскладного речення утворюють підмет і присудок. Двоскладні речення подані в кожній мові кількома структурними схемами залежно від граматичного оформлення підмета і присудка і залежно від характеру зв'язку між ними [97]. Нижче розглянуто основні групи речень УСМ і наведено дерева синтаксичного розбору для кожної з них.

3.2.1. Правила перекладу речень української словесної мови на українську жестову

Стверджувальні речення. Дерево синтаксичного розбору для речення УСМ «Я дивлюсь цікавий фільм» зображено на рис. 3.4.



Рис. 3.4. Дерево синтаксичного розбору речення
«Я дивлюсь цікавий фільм»

Оскільки українська словесна мова є флективною, то це саме речення ми можемо побудувати інакше: «Я цікавий фільм дивлюсь» або «Я цікавий дивлюсь фільм». Зміст речення у такому разі не зміниться. Такого не можна стверджувати щодо української жестової мови, в якій послідовність слів чітко визначена та підпорядковується правилам граматики жестової мови. Найчастіше слова в українській жестовій розташовуються у порядку SVO («Subject + Verb + Object»), що є прямим порядком розташування слів. Отже, враховуючи особливості української жестової мови, вищенаведене речення на УЖМ буде відтворене так: «Я дивитися фільм цікавий».

На основі правил граматики української жестової мови визначено правила перекладу, за допомогою яких отримаємо речення УЖМ за допомогою дерева синтаксичного розбору речення УСМ.

Правило № 1. Якщо потрібно відтворити розповідне стверджувальне речення на УЖМ, тоді першим відтворюється суб'єкт (підмет), потім предикат (присудок) і об'єкт (додаток). Після додатка відтворюється його ознака (прикметник). Це саме стосується і ознаки підмета. Отже, набір правил синтаксису для стверджувального речення УЖМ матиме такий вигляд:

VERB->show (Subject, *, Object),

SUBJECT->show (*, Adj1, Adj2,..., AdjN),

OBJECT->show (*, Adj1, Adj2,..., AdjN).

Позначення часу в УЖМ. Для перекладу речення, важливо знати, який час мають основні дієслова. Якщо йдеться про теперішній час – переклад здійснюється без будь-яких уточнень часу. Якщо ж дієслова в минулому чи майбутньому часі, переклад ведеться з додаванням часових жестів, тому що у жестовій мові для позначення дії предмета вживається інфінітив, а також допоміжні дієслова (Verb_a, auxiliary verb). Наприклад, «читав» ми перекладемо як «читати + був», а «читатиму» як «читати + буду».

Це розглянуто на прикладі речення УСМ «Я подивилася цікавий фільм» (рис. 3.5). У цьому реченні дієслово знаходиться у минулому часі.

Правило 2. Для позначення часу в УЖМ використовуються допоміжні жести «буду» (майбутній), «був», «вже» (минулий), які додаються до основного жесту. Так, уточнені правила синтаксису для минулого та майбутнього часу матимуть такий вигляд:

VERB[FUTURE]->show (Subject, *, буду, Object),

VERB[PAST]->show (Subject, *, був, Object).

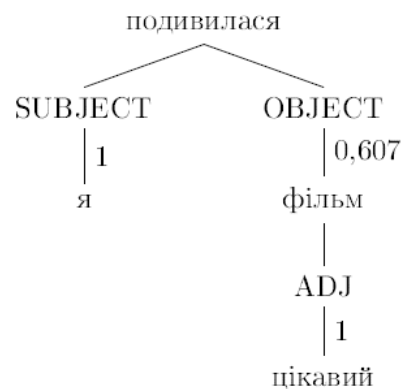


Рис. 3.5. Дерево синтаксичного розбору речення

«Я подивилася цікавий фільм»

Правило № 3. Під час побудови розповідного стверджувального речення прислівники використовуються після тих слів, до яких вони стосуються. Отже, набір правил синтаксису для наведеного речення УЖМ матиме такий вигляд:

VERB->show (Subject, *, ADV, Object),

VERB[FUTURE]->show (Subject, *, буду, ADV, Object),

VERB[PAST]->show (Subject, *, був, ADV, Object).

Розглянено дерево синтаксичного розбору для речення УСМ «Я нарешті прочитав цей текст» (рис. 3.5).

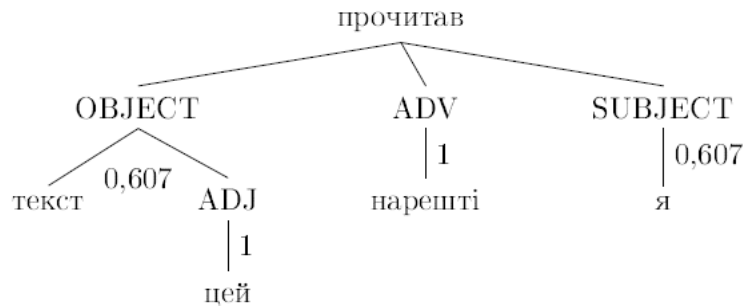


Рис. 3.6. Дерево синтаксичного розбору речення
«Я нарешті прочитав цей текст»

Заперечні речення. В УЖМ заперечні одиниці «ні», «не буде», «ніколи», «не хочу» тощо, які є основою заперечних жестових конструкцій, застосовуються лише після жестів, які вони заперечують. Розглянемо це на прикладі речення УСМ «Я ніколи не читав цей текст» (рис. 3.7).

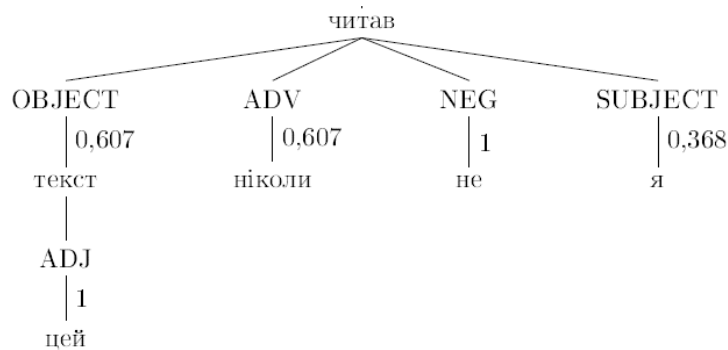


Рис. 3.7. Дерево синтаксичного розбору речення
«Я ніколи не читав цей текст»

Правило № 4. Для відтворення розповідного заперечного речення, заперечні жести «ні», «не буде», «ніколи», «не хочу», використовуються після тих жестів, які вони заперечують. Отже, набір правил синтаксису для наведеного речення УЖМ матиме такий вигляд:

VERB[NEG]->show (Subject, *, не, ADV, Object),

VERB[FUTURE, NEG]->show (Subject, *, буду, не, ADV, Object),

VERB[PAST, NEG]->show (Subject, *, був, не, ADV, Object).

Питальні речення. В українській жестовій мові для побудови питального речення застосовується таке правило: питальні слова (наприклад, «як», «коли», «чому», «скільки»), завжди ставляться в кінці речення.

Питальні слова можуть належати до різних частин мови, тому для якісного перекладу класифікуємо їх на такі групи: питальні займенники («хто», «що», «який», «чий», «котрий», «скільки», «де», «куди», «звідки», «як», «коли», «доки», «чому», «навіщо» та ін.), питальні частки («чи», «хіба», «невже»)[97]. Крім того, враховано типи питальних речень:

- 1) загальне питання – до всього речення (наприклад, речення «Ти студент?»);
- 2) спеціальні питання – до другорядних членів речення (наприклад, «Де ти працюєш?»).

Для того, щоб описати правила трансформації питального речення з УСМ на УЖМ наведено дерево синтаксичного розбору для речення УСМ «Коли ти купив машину?» (рис. 3.8):

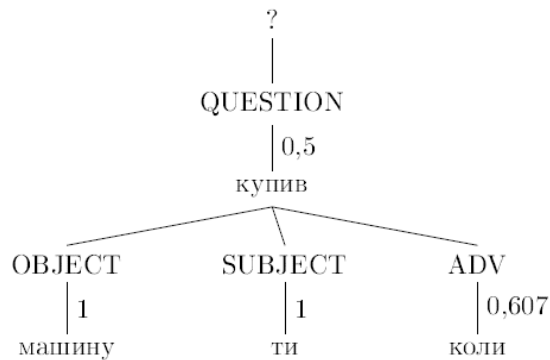


Рис. 3.8. Дерево синтаксичного розбору речення «Коли ти купив машину?»

Правило № 5. Якщо потрібно відтворити питальне речення на УЖМ, тоді першим відтворюється суб'єкт, потім предикат, об'єкт і питальне слово. Для показу питального речення у жестовій мові використовується питальна міміка обличчя (*question face expression, QFE*) (у словесній мові застосовується знак питання). Оскільки речення жестовою мовою ми записуємо, то і знак

питання використовуватимемо. Отже, набір правил синтаксису для питального речення УЖМ матиме такий вигляд:

- питальне слово – займенник (Pron, pronomen):

*VERB[QUESTION]->show (QFE, Subject, * , Object, Pron),*

*VERB[FUTURE, QUESTION]->show (QFE, Subject, * , буду, Object, Pron),*

*VERB[PAST, QUESTION]->show (QFE, Subject, * , був, Object, Pron).*

- питальне слово – частка (Pt, particle):

*VERB[QUESTION]->show (QFE, Subject, * , Object, Pt),*

*VERB[FUTURE, QUESTION]->show (QFE, Subject, * , буду, Object, Pt),*

*VERB[PAST, QUESTION]->show (QFE, Subject, * , був, Object, Pt).*

Окличні речення. Як в українській словесній мові, так і у жестовій окличні речення можуть виражати: здивування, іронію, розгубленість, сумнів, невпевненість, упевненість, жах, злість, гнів, жаль, ніжність, любов, захоплення, докір, осуд, погрозу, наказ, заклик, прохання та ін. [29]. Ці речення вимовляються підвищеним тоном з особливою інтонацією, в кінці яких в УСМ ставиться знак оклику. В українській жестовій мові окличні речення передаються за допомогою виразу обличчя (міміки) чи сплеску руками, тобто будь-яким природним жестом.

Синтаксичний розбір окличного речення УСМ «Я нарешті прочитав цей текст!» наведено на рис. 3.9.

Правило № 6. Для відтворення окличного речення на УЖМ, першим відтворюється суб'єкт, потім предикат, об'єкт. Для показу окличного речення у жестовій мові використовується міміка обличчя (*exclamation face expression, EFE*) (у словесній мові застосовується знак оклику). Оскільки речення жестовою мовою ми записуємо, то і знак оклику використовуватимемо. Отже, набір правил синтаксису для окличного речення УЖМ матиме такий вигляд:

*VERB[EXCLAMATION]->show (EFE, Subject, * , ADV, Object),*

*VERB[FUTURE, EXCLAMATION]->show (EFE, Subject, * , буду, ADV, Object),*

VERB[PAST, EXCLAMATION]->show (EFE, Subject, *, був, ADV, Object).

Для тестування перекладу було взято прості речення з корпусу паралельних текстів «Українська словесна мова – Українська жестова мова», який містить 220 речень.

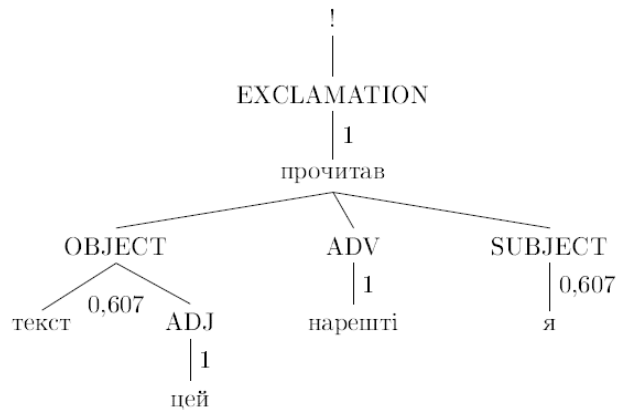


Рис. 3.9. Дерево синтаксичного розбору речення
«Я нарешті прочитав цей текст!»

Блок-схема алгоритму перекладу української словесної мови на анотовану українську жестову на основі правил зображена на рис. 3.10.

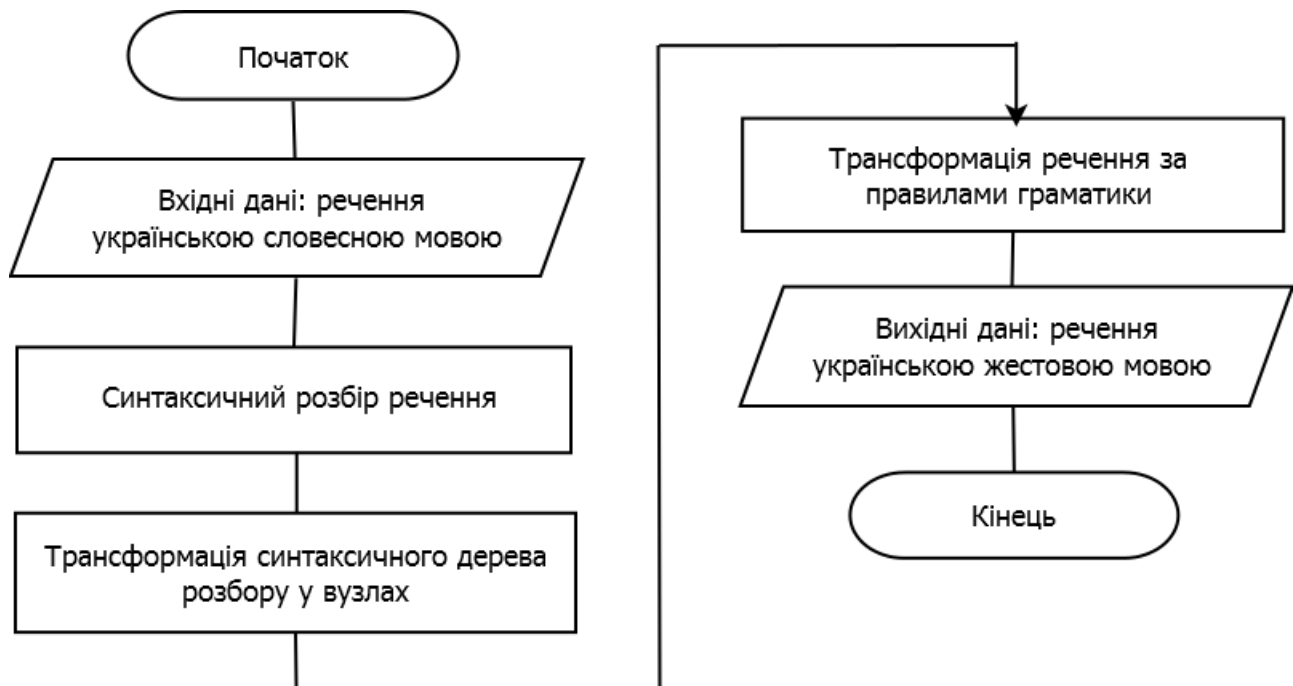


Рис. 3.10. Блок-схема алгоритму перекладу української словесної мови на анотовану українську жестову на основі правил

Під час перекладу з української словесної на анотовану українську жестову мову враховувалось, що усі іменники, займенники, прикметники та числівники в українській жестовій мові не мають закінчень, за якими можна ідентифікувати відмінок, рід та число, як у словесній мові. Також в УЖМ немає прийменників, а напрямок руху та розміщення жестів вже закладені у самих жестах. В українській словесній мові за закінченнями дієслів визначаються число, особа, рід, час, вид та спосіб, а у жестовій мові всі дієслова є у неозначеній формі дієслова.

Проте, враховуючи описані вище правила, під час перекладу виникали помилки. Нижче наведено кілька прикладів машинного перекладу з УСМ на УЖМ (табл. 3.3):

Таблиця 3.3

Приклади перекладу УСМ→УЖМ

№	Українська словесна мова	Українська жестова мова (машинний переклад)	Українська жестова мова
1	Дорога йде прямо.	Дорога йти прямо.	Дорога прямо.
2	Серце б'ється швидко.	Серце битися швидко.	Серцебиття (швидкий темп).
3	Як тебе звати?	Ти називати як?	Ти ім'я?

У першому, другому та третьому реченнях помилки виникли через те, що одному жесту може відповідати декілька слів. Тому для правильного генерування перекладу треба створити словник таких жестів та їх відповідних перекладів українською словесною мовою. Крім того, відповідно до контексту речення, в УЖМ використовуються різні жести на позначення дії. Наприклад, дієслово «йти» у різних контекстах позначається різними жестами: «я йду» і «час йде» (рис. 3.11).

Для вирішення цього завдання потрібно створити двомовний словник концептів та правил визначення, до якого концепту належить слово у певному контексті для української жестової мови.



а



б

Рис. 3.11. Зображення жестів: а – «я йду»; б – «час йде»

3.2.2. Правила перекладу речень української жестової мови на українську словесну

На граматики української жестової мови визначено правила перекладу, за допомогою яких отримуємо речення УСМ.

Правило № 1. Якщо потрібно відтворити розповідне стверджувальне речення на УСМ, тоді першим відтворюється суб'єкт (підмет), потім предикат (присудок) і об'єкт (додаток). Ознака (прикметник) відтворюється перед означуваним словом (підметом чи додатком). Отже, набір правил синтаксису для стверджувального речення УСМ матиме такий вигляд:

VERB->show (Subject, *, Object),

SUBJECT->show (Adj1, Adj2,..., AdjN, *),

OBJECT->show (Adj1, Adj2,..., AdjN, *).

Позначення часу в УЖМ. В українській жестовій мові для позначення дії предмета вживається інфінітив, а також допоміжні дієслова для позначення минулого та майбутнього часів (Verb_a). Наприклад, «читав» ми перекладемо як «читати + був», а «читатиму» як «читати + буду». Для дієслів теперішнього часу – переклад здійснюється без будь-яких уточнень часу. Це розглянено на прикладі речення УЖМ «Я читати був книга цікавий». У цьому реченні дієслово знаходиться у минулому часі. Отже, переклад на українську словесну мову цього речення такий: «Я читав цікаву книгу».

Правило № 2. Якщо у реченні УЖМ використовуються допоміжні жести «буду» (майбутній), «був», «вже» (минулий), які додаються до основного жесту, тоді у перекладеному реченні УСМ дієслова будуть у минулому або майбутньому часі без допоміжних дієслів. Так, уточнені правила синтаксису для минулого та майбутнього часу матимуть такий вигляд:

VERB[FUTURE]->show (Subject, * , Object),

VERB[PAST]->show (Subject, * , Object).

Правило № 3. Під час побудови розповідного стверджувального речення обставини (прислівники) використовуються або перед, або після тих слів, до яких вони стосуються. Отже, набір правил синтаксису для наведеного речення УСМ матиме такий вигляд:

1) якщо обставина у реченні використовується перед присудком:

VERB->show (Subject, ADV, * , Object),

VERB[FUTURE]->show (Subject, ADV, * , Object),

VERB[PAST]->show (Subject, ADV, * , Object);

2) якщо обставина у реченні використовується після присудка:

VERB->show (Subject, * , ADV, Object),

VERB[FUTURE]->show (Subject, * , ADV, Object),

VERB[PAST]->show (Subject, * , ADV, Object).

Для прикладу розглянено речення жестовою мовою «Я знати добре жестова мова». Оскільки в українській словесній мові, порівняно із жестовою, послідовність слів є довільною, то це речення в УСМ можна побудувати так, як подано у табл. 3.4.

Таблиця 3.4

Приклади перекладу з УЖМ на УСМ

УЖМ	УСМ
Я знати добре жестова мова	1) Я добре знаю жестову мову
	2) Я знаю добре жестову мову

Заперечні речення. В УСМ заперечні одиниці «ні», «не буде», «ніколи», «не хочу» тощо, на відміну від УЖМ, застосовуються перед тими словами, які вони заперечують.

Правило № 4. Для відтворення розповідного заперечного речення, заперечні жести «ні», «не буде», «ніколи», «не хочу», використовуються перед тими словами, які вони заперечують. Отже, набір правил синтаксису для заперечних речень УСМ матиме такий вигляд:

VERB[NEG]->show (Subject, ADV, не, *, Object),

VERB[FUTURE, NEG]->show (Subject, ADV, не, *, Object),

VERB[PAST, NEG]->show (Subject, ADV, не, *, Object).

Застосування цього правила розглянено на прикладі речення УЖМ «Я вчити не уроки». Відповідно до правила, наведене речення українською словесною мовою: «Я не вчу уроки». Крім того, одиниці СМ, які містять заперечення, в УЖМ, за можливості, замінюються не відповідними, а іншими жестами. Це стосується прикметників із заперечною часткою «не». Наприклад, «не дорогий» – «дешевий», «не весело» – «сумно». Для вирішення цієї задачі потрібно використовувати словник антонімів.

Питальні речення. Одним із засобів вираження питального речення в українській словесній мові є вживання питальних займенників («хто», «що», «який», «чий», «котрий», «скільки», «де», «куди», «звідки», «як», «коли», «доки», «чому», «навіщо» та ін.), питальних часток («чи», «хіба», «невже»), які, будучи членом речення, підсилюють питальне значення цього речення. Питальне слово в них, як правило, стоїть на початку [97]. В українській жестовій мові для побудови питального речення застосовується таке правило: питальні слова завжди ставляться в кінці речення.

Крім того, для правильного генерування перекладу враховано типи питальних речень:

1) загальне питання – до всього речення (наприклад, речення «Ви вчитель?»);

2) спеціальні питання – до другорядних членів речення (наприклад, «Де ти живеш?»).

Для того, щоб описати правила трансформації питального речення з УЖМ на УСМ наведено приклади перекладу речень (табл. 3.5).

Таблиця 3.5

Приклади перекладу з УЖМ на УСМ

УЖМ	УСМ
1) Вчитель ви?	1) Ви вчитель?
2) Ти жити де?	2) Де ти живеш?

Правило № 5. Під час відтворення питального речення на УСМ першим відтворюється питальне слово, потім суб'єкт, предикат і об'єкт. Для показу питального речення у словесній мові застосовується знак питання, у жестовій мові використовується питальна міміка обличчя (*question face expression, QFE*). Отже, набір правил синтаксису для питального речення УСМ матиме такий вигляд:

1) питальне слово – займенник (Pron, pronomen):

*VERB[QUESTION]->show (Subject (Pron), * , Object, «?»),*

*VERB[FUTURE, QUESTION]->show (Subject (Pron), * , Object, «?»),*

*VERB[PAST, QUESTION]->show (Subject (Pron), * , Object, «?»);*

2) питальне слово – частка (Pt, particle):

*VERB[QUESTION]->show (Pt, Subject, * , Object, «?»),*

*VERB[FUTURE, QUESTION]->show (Pt, Subject, * , Object, «?»),*

*VERB[PAST, QUESTION]->show (Pt, Subject, * , Object, «?»).*

Окличні речення. В українській жестовій мові окличні речення передаються за допомогою виразу обличчя (міміки) та жестикуляції (сплеск руками), тобто будь-яким природним жестом.

Правило № 6. Для перекладу окличного речення на УСМ першим відтворюється суб'єкт, потім предикат, об'єкт. Для показу окличного речення у словесній мові застосовують знак оклику, у жестовій мові використовують

міміку обличчя (*exclamation face expression, EFE*). Отже, набір правил синтаксису для окличного речення УСМ матиме такий вигляд:

VERB[EXCLAMATION]->show (Subject, * , ADV, Object, «!»),

VERB[FUTURE, EXCLAMATION]->show (Subject, ADV, * , Object, «!»),

VERB[PAST, EXCLAMATION]->show (Subject, ADV, * , Object, «!»).

Для тестування машинного перекладу на основі правил з української жестової мови на українську словесну використано корпус паралельних текстів «Українська словесна мова – Українська жестова мова».

Блок-схема алгоритму перекладу анотованої української жестової мови на українську словесну на основі правил зображена на рис. 3.12.

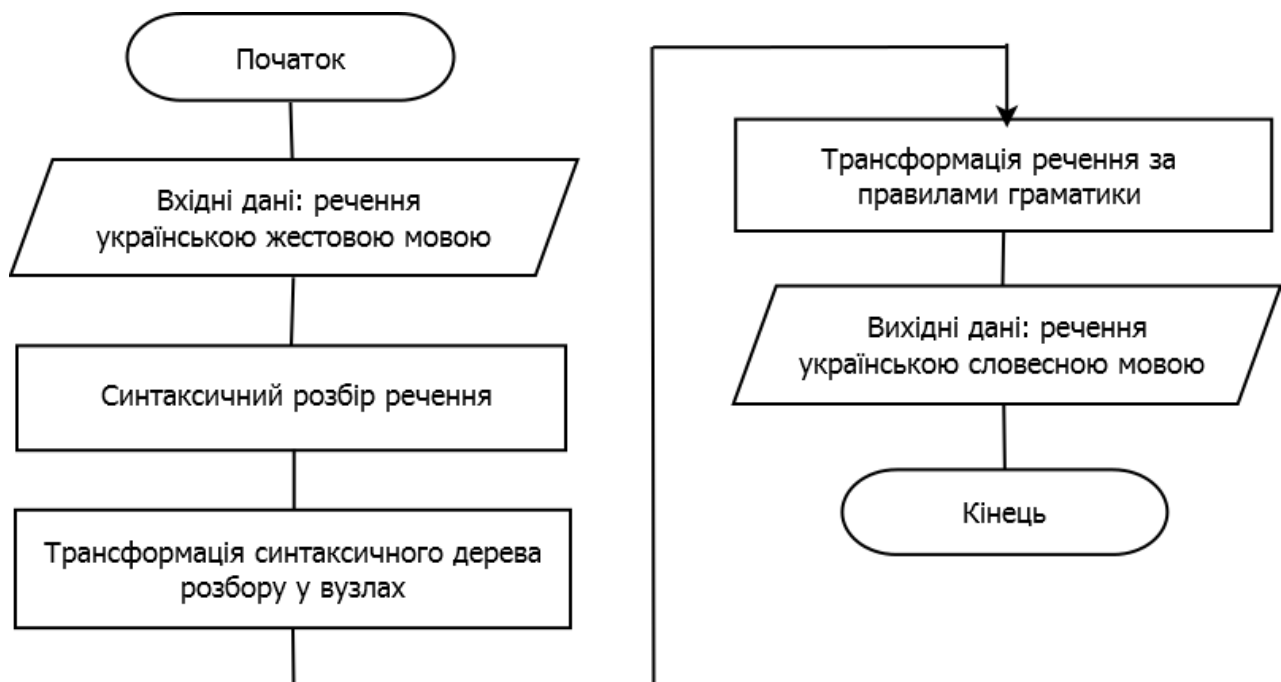


Рис. 3.12. Блок-схема алгоритму перекладу анотованої української жестової мови на українську словесну на основі правил

Основними помилками, які виникали під час перекладу з УЖМ на УСМ є невідповідність закінчень іменників, займенників, прикметників та дієслів. Це пов'язано з тим, що усі іменники, займенники та прикметники в українській жестовій мові мають не мають закінчень, по яких можна ідентифікувати

відмінок, рід та число, як у словесній мові. Також в УЖМ немає прийменників, а напрямок руху та розміщення жестів вже закладені у самих жестах.

В українській словесній мові за закінченням дієслів визначають число, особа, рід, час, вид та спосіб, а у жестовій мові всі дієслова подаються у неозначеній формі дієслова. Тобто, для якісного перекладу потрібно за допомогою супутніх жестів визначити усі морфологічні ознаки іменників, займенників, прикметників та дієслів. Для цього потрібно деталізувати правила перекладу з УЖМ на УСМ, враховуючи граматичні особливості української жестової мови.

Також під час перекладу потрібно враховувати такі додаткові правила:

- один жест може відповідати одному слову;
- унікальний жест можуть утворювати декілька слів;
- кілька жестів може відповідати одному слову;
- генерування декількох жестів із декількох слів із урахуванням певних зв'язків між ними.

Проте, враховуючи вищеописані правила, під час перекладу все ж виникали помилки. Нижче наведено кілька прикладів машинного перекладу з УЖМ на УСМ та відповідний переклад у корпусі паралельних текстів цих речень (табл. 3.6).

Таблиця 3.6

Приклади перекладу УЖМ→УСМ

№	Українська жестова мова	Українська жестова мова (машинний переклад)	Українська словесна мова
1	Сестра ти скільки?	Скільки ти сестра?	Скільки в тебе є сестер?
2	Він любити лижі.	Він любить лижі.	Він любить кататися на лижах.
3	Діти ти?	Твої діти?	У тебе є діти?
4	Дорога рівно.	Дорога рівно.	Дорога йде рівно.

3.3. Застосування методу машинного перекладу анотованої української жестової мови з використанням ГДО

Відомі підходи перекладу жестових мов з використанням статистичних методів перекладу [49], методів на основі правил [94] та онтологій [50] не забезпечують необхідної якості перекладу, що зумовлює необхідність використання альтернативного підходу з використанням граматично доповненої онтології.

Блок-схему алгоритму перекладу на українську жестову мову із застосуванням граматично доповненої онтології зображено на рис. 3.13.

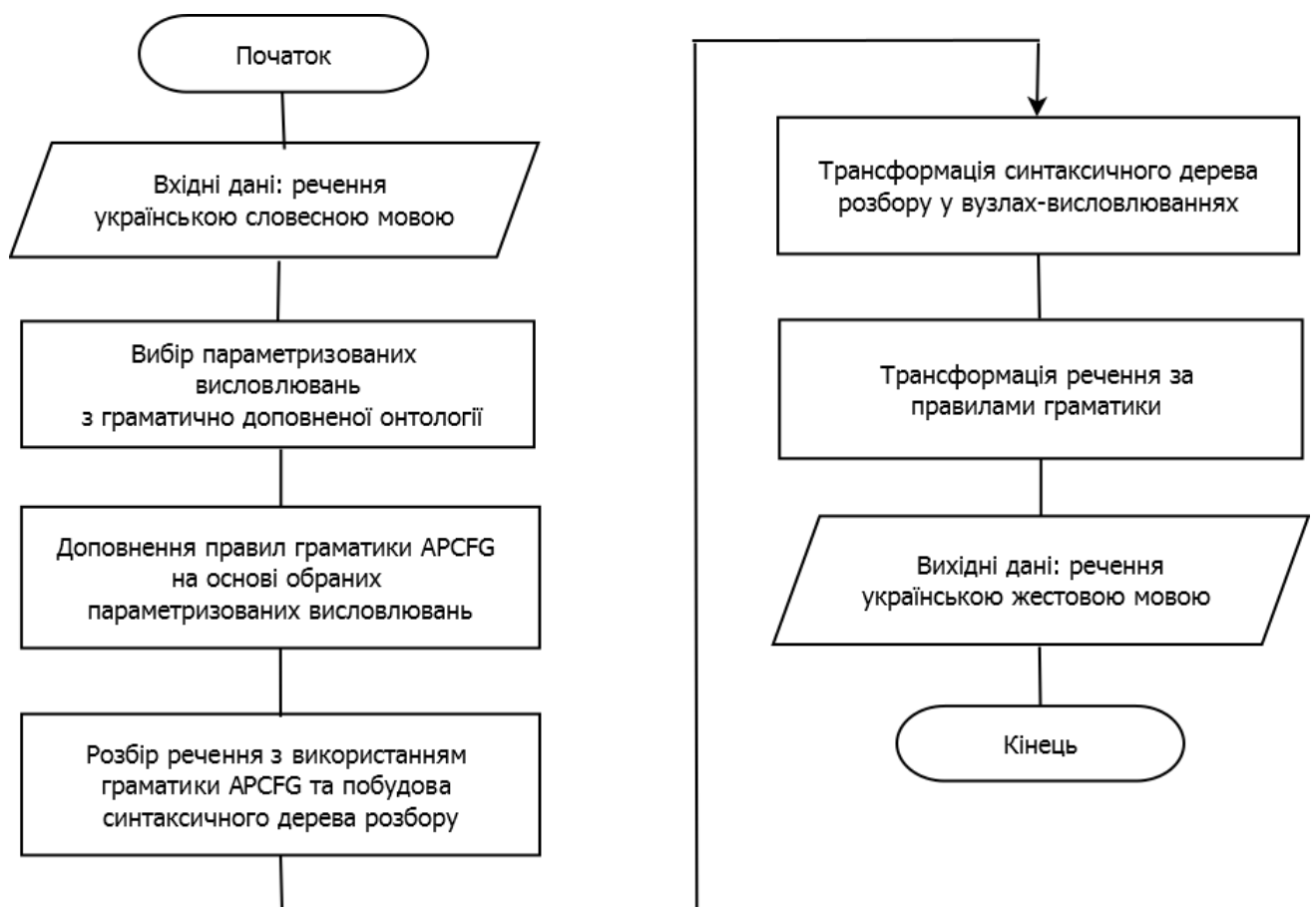


Рис. 3.13. Блок-схема алгоритму перекладу із застосуванням граматично доповненої онтології

На вхід подається речення українською словесною мовою. Далі програма здійснює пошук параметризованих висловлювань 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 речення, коли розбір відповідає граматично доповненій онтології (вага правила більша за 1).

Після семантично-синтаксичного розбору речення виконується трансформація цього речення за заданими правилами граматики УЖМ. На цьому етапі здійснюється встановлення визначеного відповідно до УЖМ порядку 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 УЖМ. Усі правила, за якими здійснюється семантично-синтаксичний розбір речень, поділено на п'ять груп. Це група правил для:

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

Перша група правил містить граматичні конструкції для побудови речень із врахуванням порядку слів у них:

END -> <.>;
FULLS -> START S END;
SHORTS -> START NP END;
FULLQ -> START NP <?>;
FULLQ -> START VP <?>;
FULLQ -> START QS <?>;
FULLQ -> START AKSTOSAY QS <?>;
GENCOMMA -> <,> / <.> / <:> / <?> / <!> / START;
S *-> S ADDRESS;
S *-> S PLACE;
S *-> S ADDITIONAL;
S *-> S OBJECT;
S *-> S FROM;
S *-> S TARGET;
S *-> S TIME;
S *-> S NAME,

де END – це крапка, FULLS – повне речення, SHORTS – неповне речення, START – початок речення, S – речення, FULLQ – повне питальне речення, NP (Noun Phrase) – іменникове сполучення, VP (Verb Phrase) – дієслівне сполучення, QS – питальне речення, AKSTOSAY – <розкажи>, <скажи>, <повідом>, <повтори>, GENCOMMA – розділовий знак.

Усі групи правил містять правила синтаксичного розбору речень. Друга, третя та четверта групи правил доповнюються правилами семантичного розбору із використанням граматично доповненої онтології.

Друга група правил для іменникових синсетів містить граматичні конструкції, в яких вживаються іменники, займенники та прикметники. Наведемо основні правила:

AN[NCG] -> noun[NCG] noun[NCGu]?; // хлопець Іван
COMMEDADJG[NCG] -> <,> adj[NCG] GENCOMMA);
COMMEDADJG[NCG] -> COMMEDADJG[NCG] <,> adj[NCG] GENCOMMA;
ADJG[NCG] -> adj[NCG];
ADJQ[NCG] -> adj[NCGq]; // якою? котрею? ...
ADJG[NCG] -> COMMEDADJG[NCG];
DNP[NCG] -> NG[NCG c2c3c4c5c6c7];
NP[NCG p3] -> NG[NCG c1];
NP[NCGP] -> pronoun[NCGP c1],

де NG – іменникова група (noun group), AN, ADJG – група прикметників, відокремлені комою, adj – прикметник, pronadj, DNP, pronoun – займенник, NG[NCG c2c3c4c5c6c7] (додаткові позначення наведено в табл. 3.7) – додатки, NG[NCG c1] – підмет, виражений іменником, pronoun[NCGP c1] – підмет, виражений займенником.

Додаткові правила для іменникових синсетів:

NG[NCG] -> adj[NCG] NG[NCG]; // «гарна дівчина»
NG[NCG] -> adj[NCG]? AN[NCG] NG[c2]? ADJG[NCG]?; // «жовта іграшка дитини, нова, яскрава»
NG[NCG] -> adj[NCG]? pronoun[NCG] ADJG[NCG]?; // «веселого тебе, здорового»

NG[C n] -> NG[C] conj NG[C]; // «брат і сестра»*
NP[NCGP] -> NP <чи> NP[NCGP]; // «брат чи сестра»

Таблиця 3.7

Позначення для правил семантично-синтаксичного розбору

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

Також описано правила додатків у реченні:

V -> у | в;

Z -> з | із | зі;

PARTP -> це | ось | оце | он | ген | осьде | то | ото

TARGET -> V DNP[c4] | <до> DNP[c2] | <додому> | <туди> | <сюди>;

NAME -> noun[NCGu]; // власна назва

ADDRESS -> DNP[c3]; //адресат дії (кому)

PLACE -> V DNP[c6] | <тут> | <там>;

ADDITIONAL -> Z DNP[c5]; // «з мамою»

TIME -> <зараз> | <потім>;

FROM -> Z DNP[c2]; //звідки

OBJECT -> DNP[c4],

де PARTP – вказівні частки.

Крім того, ця група правил доповнена правилами з граматично доповненої онтології, які позначають відношення між поняттями (відношення гіперонімії/гопонімії, голонімії/меронімії). Наприклад, поняття «людина» є гіперонімом для поняття «дитина»: *<людина>[NCG] -> <дитина>[NCG]*.

Аналогічне відношення гіперонімії/гопонімії існує між поняттями «історія-розповідь» та «казка», «оповідання», «пригода»:

<історія-розповідь>[NCG] -> <казка>[NCG];

<історія-розповідь>[NCG] -> <оповідання>[NCG];

<історія-розповідь>[NCG] -> <пригода>[NCG].

Третя група правил містить основні правила для синтаксичного розбору дієслівних синсетів:

*VP *-> VP ADDRESS;*

*VP *-> VP PLACE;*

*VP *-> VP ADDITIONAL;*

*VP *-> VP OBJECT;*

*VP *-> VP FROM;*

*VP *-> VP TARGET;*

*VP *-> VP TIME;*

P -> VP adv;

VP[PNM] -> verb[PNM];

VP[PN m-] 1.1-> verb[PNm+] IVP;

VP[p1p2p3p-N] -> ADJG[N]; // прикметник в ролі дієслова

VP[PNM] -> <ne> VP[PNM],

де *adv* – прислівник, *verb[PNm+]* – модальне дієслово, *IVP* – інфінітив, *<ne>* - заперечення.

Також описано додаткові правила для дієслівних синсетів, що містять інфінітив. Наприклад, для речення «Я піду в далекі гори» під час розбору буде використано правило «*IVP *-> IVP PLACE*». Перелік додаткових правил подано нижче:

IVP -> <ne>? verb[i];

*IVP *-> IVP ADDRESS;*

*IVP *-> IVP PLACE;*

*IVP *-> IVP ADDITIONAL;*

*IVP *-> IVP OBJECT;*

*IVP *-> IVP FROM;*

*IVP *-> IVP TARGET;*

*IVP *-> IVP TIME;*

*IVP *-> IVP NAME;*

Для дієслівних синсетів, окрім правил синтаксичного розбору, описані онтологічні правила, які позначають відношення гіперонімії/гопонімії, імлікації

та правила для вираження граматичних конструкцій дієслівних синсетів. Наприклад, для дієслівного синсету «розповідати» вказано граматичну конструкцію для вираження:

*<розповідати-синсет>[PNM] * 1.1-> <розповідати>[PNM] <історія-розповідь>[c4] <слухач>[c3]?");*

<verb[PNM] -> <розповідати-синсет>[PNM].

Четверта група правил описує граматичні конструкції питальних речень, наприклад:

QS -> S;

QS -> <чи> S;

QS -> pronoun[NGC] DNP[NGC] DNP? S?;

QS -> adj[NGCq] DNP[NGC] DNP? S?;

QS -> ADJQ[NGC] DNP[NGC] DNP? S?;

QS -> <чи> QS;

*QS *-> adv S;*

*QS *-> adv IVP;*

*QS *-> pronoun[q] NP;*

QS -> pronoun[q] S;

ADVQ -> де | як | коли | скільки | звідки,

де ADVQ – питальний прислівник.

Для семантично-синтаксичного розбору речень використовують **п'яту групу правил** для речень, у яких порушено порядок слів. Наприклад, для речення «Додому я йду» використовується правило «S *-> TARGET S».

*S *-> ADDRESS S;*

*S *-> PLACE S;*

*S *-> ADDITIONAL S;*

*S *-> OBJECT S;*

*S *-> FROM S;*

*S *-> TARGET S;*

*S *-> TIME S;*

*S *-> NAME S.*

Крім того, розроблено правила семантично-синтаксичного розбору речень української жестової мови. Ці правила поділяються на чотири групи та аналогічні до перших чотирьох груп української словесної мови. *Перша група правил* містить граматичні конструкції для побудови жестових речень із урахуванням послідовності слів у них, і ці правила такі самі, як і для УСМ.

Друга група правил для іменникових синсетів української жестової мови відрізняється від правил для іменникових синсетів УСМ.

NG -> AN;

NG -> NG ADJG;

NG 0.5-> ADJG NG; // означення рідко вживається до іменника

NG -> pronadj NG; // мій, твій навпаки частіше вживаються до

іменника

NG 0.5 -> NG pronadj;

DNP -> NG;

NP -> NG;

NP -> pronoun;

DNP -> pronoun;

NG[n] -> NG NG;*

NG[n] -> num NG;*

ADJG -> adj,

AN -> noun noun[u]?; // хлопець Петро

// щоб не робити категорію для всіх слів

AN -> <жест> noun;

AN -> <жест> verb;

AN -> <жест> adj,

де n* – іменник у множині (множина у жестовій мові виражається повторенням слова або вказанням кількості num), u – власна назва.

Правила додатків у реченні УЖМ аналогічні до правил додатків у реченні УСМ.

Крім того, ця група правил для речень УЖМ також доповнена відношеннями гіперонімії/гопонімії, голонімії/меронімії.

Третя група правил для дієслівних синсетів УЖМ містить правила такі самі, як і для УСМ. Додаткові правила для дієслівних синсетів УЖМ такі:

VP -> verb / adjpart;

VP[p1p2p3 N] -> <ε> adj[N c4c5]; // TARGET?

VP[p1p2p3 N] -> <ε> DNP[N c4c5];

S[n1] -> <ву> <ε> DNP[n1 c4c5];

Четверта група правил для розбору жестових речень аналогічна до четвертої групи правил розбору українських словесних речень.

Розглянемо побудову дерев семантично-синтаксичного розбору для речень української словесної та жестової мов та правила розбору для них. Наприклад, для речення УСМ «Я розповідаю дітям казку» дерево семантично-синтаксичного розбору зображено на рис. 3.14.

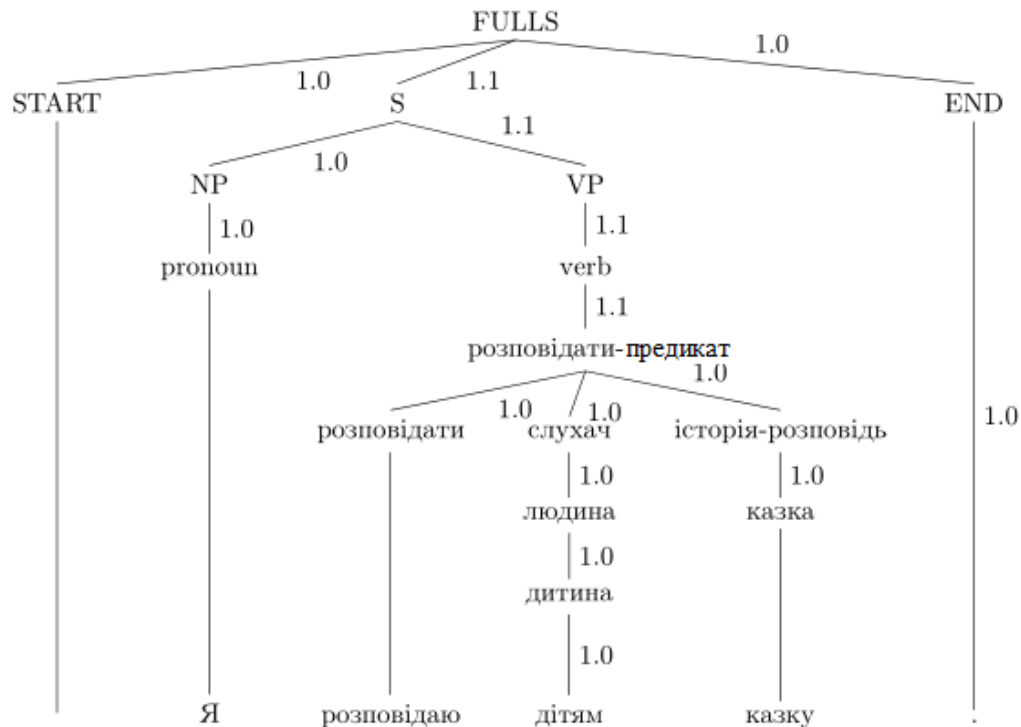


Рис. 3.14. Дерево семантично-синтаксичного розбору для речення «Я розповідаю дітям казку»

Правила семантики, які створені з онтології для дерева семантично-синтаксичного розбору української словесної мови, зображені на рис. 3.14:

<людина>[NCG] → <дитина>[NCG]

<слухач>[NCG] → <людина>[NCG]

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

За допомогою відношень між поняттями у граматично доповненій онтології на місця параметризованих змінних можна поставити лише правильні за суттю 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. 3.15.

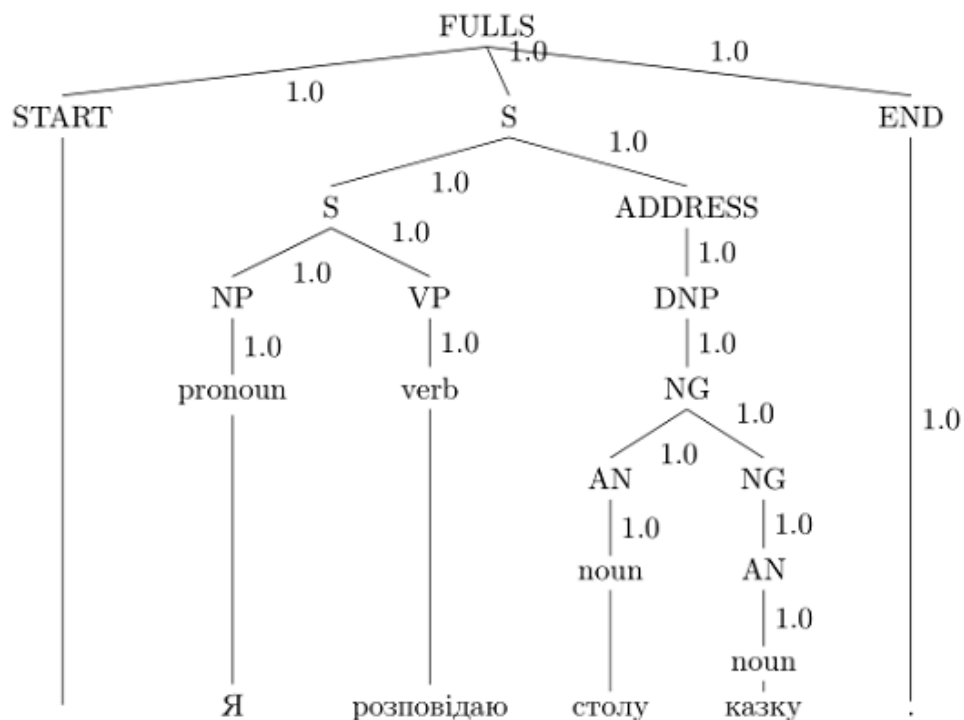


Рис. 3.15. Дерево семантично-синтаксичного розбору для речення «Я розповідаю столу казку»

Для цього дерева семантично-синтаксичного розбору речення використовуються звичайні правила синтаксису:

$S * \rightarrow S \text{ ADDRESS},$

$S * \rightarrow VP[NP] NP[NP],$

ADDRESS -> DNP[c3],

DNP[NCG] -> NG[NCG c2c3c4c5c6c7].

Оскільки у синсет включено граматичну конструкцію української жестової мови, то це дає змогу побудувати дерево семантично-синтаксичного розбору для речень УЖМ. Наприклад, для речення «Я розповідати діти казка» дерево розбору зображено на рис. 3.16. Можливість побудови дерев семантично-синтаксичного розбору для речень української словесної та жестової мов дає змогу здійснити однозначний переклад між цими мовами.

Для української жестової мови правила семантики, які створені з онтології (рис. 3.16):

<слухач>[NCG] → <дити>[NCG]

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

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

verb[PN] → <розповідати-синсет>[PN]

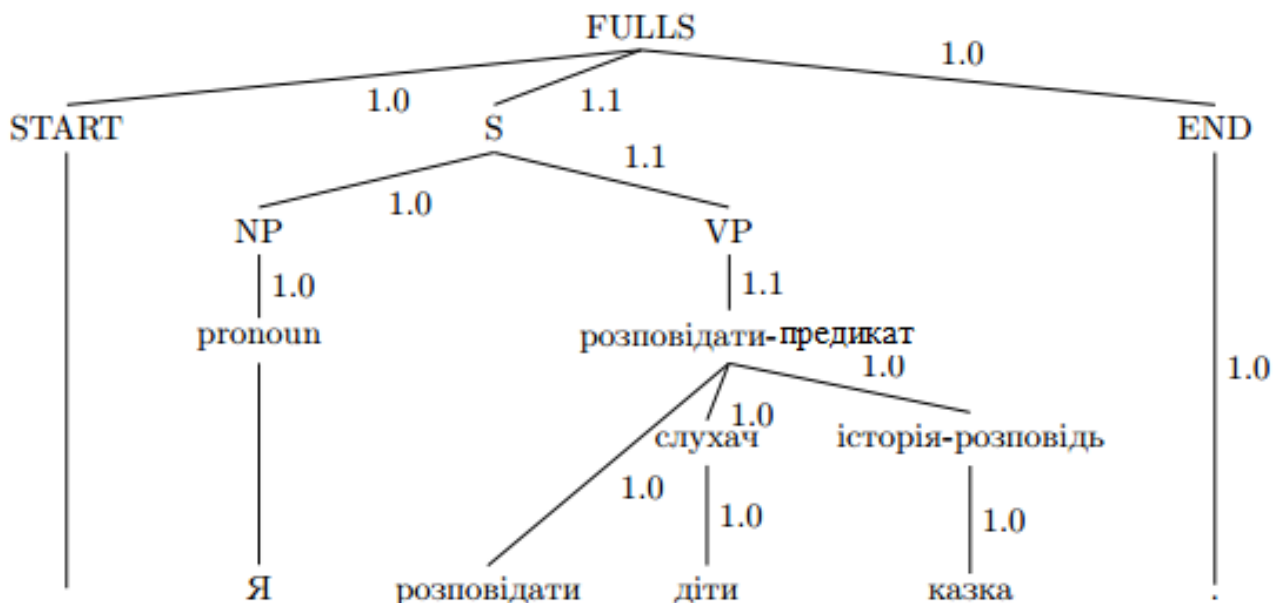


Рис. 3.16. Дерево семантично-синтаксичного розбору для речення

«Я розповідати діти казка»

У результаті проведених лінгвістичних досліджень української жестової мови описано правила обходу дерева семантично-синтаксичного розбору речень українською словесною мовою для синтезу речень анотованою українською жестовою мовою. На основі прикладів дерев семантично-синтаксичного розбору речень української словесної мови визначено основні правила перекладу на анотовану українську жестову мову.

Використання граматично доповненої онтології для семантично-синтаксичного розбору речень дало змогу здійснити правильний розбір для тих речень, у яких під час синтаксичного розбору виникали помилки. Наприклад, для речення «Моє серце б'ється швидко» дерево синтаксичного розбору зображено на рис. 3.17:

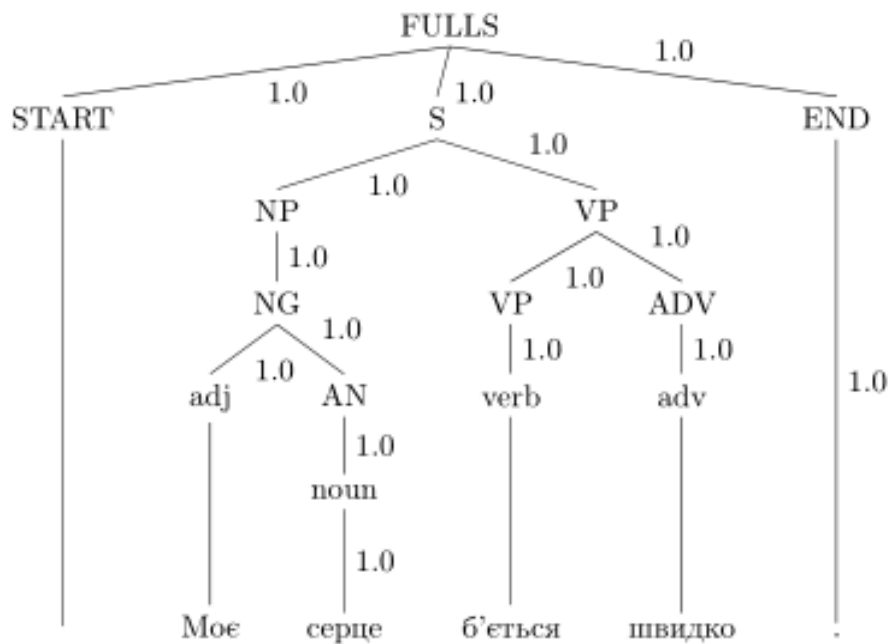


Рис. 3.17. Дерево синтаксичного розбору речення
«Моє серце б'ється швидко»

Через це здійснено неправильний переклад українською жестовою мовою («СЕРЦЕ БИТИ ШВИДКО»). У цьому разі жест «БИТИ» означає «завдавати ударів», тобто повністю втрачено зміст речення. В УЖМ словосполучення «серце б'ється швидко» показується лише одним жестом «СЕРЦЕБИТТЯ».

Ураховуючи це, дерево семантично-синтаксичного розбору для речення «Мое серце б'ється швидко» зображено на рис. 3.18.

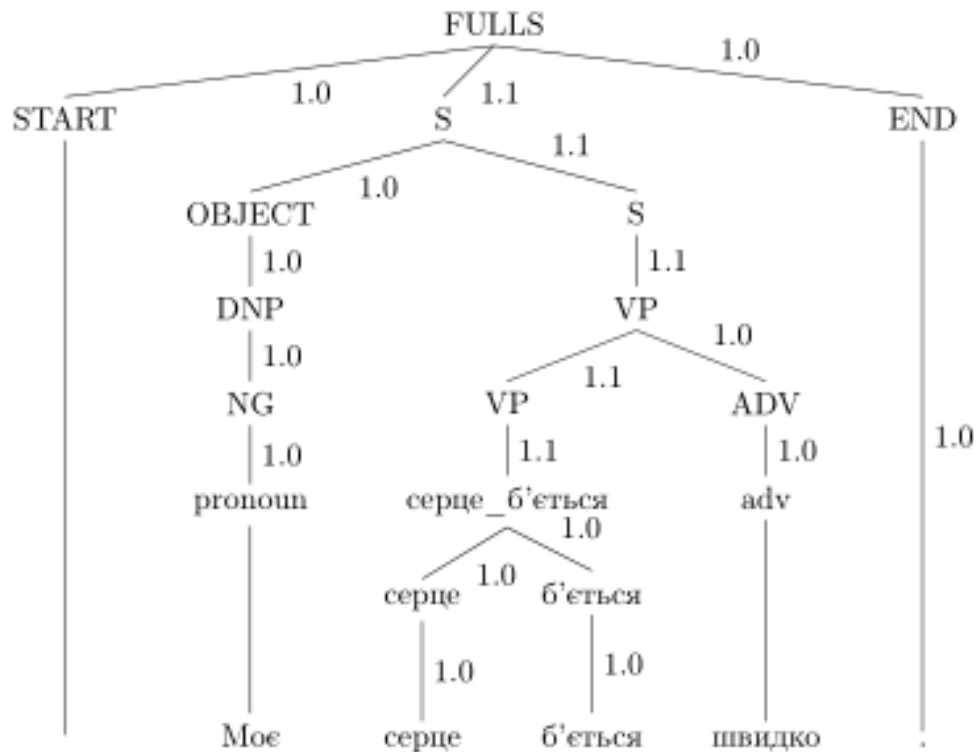


Рис. 3.18. Дерево семантично-синтаксичного розбору речення «Мое серце б'ється швидко» (з використанням ГДО)

Результати експерименту семантично-синтаксичного розбору проведені для 220 тестових речень українською словесною та жестовою мовами та наведені в табл. 3.8. Відсоток правильно проаналізованих речень з використанням ГДО є більшим, ніж під час аналізування тільки за допомогою граматичних правил.

Таблиця 3.8

Відсоток правильно синтаксично розібраних речень УСМ та УЖМ

Множина правил	Українська жестова мова	Українська словесна мова
Лише граматичні правила	72 %	65 %
Граматичні правила + + правила з ГДО	91 %	90 %

З табл. 3.8 зрозуміло, що відсоток правильно проаналізованих речень УЖМ є більшим за відсоток правильно проаналізованих речень УСМ. Це пов'язано з тим, що граматики української словесної мови має гнучкий порядок слів, а послідовність слів у жестовій мові фіксується в більшості висловлювань.

3.3.1. Трансформація дерева складових у дерево залежностей

Для перекладу речень УСМ у речення анотованої УЖМ розроблено правила трансформації дерев граматики залежностей [6, 7]. Оскільки під час семантично-синтаксичного аналізу речень УСМ будується дерево граматики складових, потрібно розробити алгоритм трансформації дерев граматики складових (семантично-синтаксичного розбору) у дерева граматики залежностей.

Для практичного втілення цього алгоритму потрібно:

- 1) описати правила визначення головної складової у дереві граматики складових;
- 2) описати послідовність кроків для трансформації дерева складових у дерева залежностей для усіх типів речень.

Алгоритм трансформації дерева складових у дерево залежностей

Після семантично-синтаксичного розбору речення виконують трансформацію дерева складових цього речення у дерево залежностей згідно з алгоритмом трансформації, поданого у [119]. Як для дерев граматики залежностей, так і для дерев граматики складових важливим є поняття «головна складова» (англ. «head»). Для дерев граматики складових головною складовою позначається головне слово у виразі і від цього слова залежать усі інші слова виразу. Ядром алгоритму є визначення головної складової кожного виразу для дерев граматики складових і встановлення зв'язку з головною складовою його батьківського вузла. Головною складовою для кожного виразу NP є вузол N в цьому вузлі NP. Головною складовою для NP є іменник (noun) або займенник (pronoun), головною складовою для VP є дієслово (verb). Головною складовою

для вузла S є головна складова VP , якщо S є простим реченням, та головна складова VP головного речення, якщо S є складним реченням.

Правила визначення головної складової в деревах граматики складових такі:

- 1) головна складова вузла VP є кореневим вузлом (корінь) у дереві граматики залежностей;
- 2) усі вузли NP та NG , у яких відсутній вузол 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 у 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 залежить від дієслова verb;
- 6) головною складовою вузлів $TARGET$, $FROM$ є додаток (DNP);
- 7) якщо вузол-нащадок S є нащадком вузла-батька S , тоді головна складовою вузла-нащадка S буде головною складовою 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». Цей алгоритм дає змогу визначити головну складову для кожного вузла дерева і відповідно залежну складову цього вузла, яка підпорядковується головній складовій. Просуваючись по дереву складових вгору, ми отримуємо головну складову всього речення (кореня дерева). На рис. 3.19 подано дерево складових і результат роботи алгоритму (чи визначення головної складової кореня дерева). Застосування алгоритму та визначення головної складової речення зображено потовщеними лініями.

Алгоритм трансформації дерева складових у дерево залежностей складається з таких кроків:

Крок 1. Шукаємо у дереві розбору вузол verb. Якщо знайшли verb, тоді переходимо до його вузла-батька VP і позначаємо verb головною складовою вузла VP. Переходимо до *кроку 2*. Якщо немає VP, тоді шукаємо вузол S і переходимо до *кроку 4*.

Крок 2. Шукаємо вузол-батько вузла VP. Якщо вузол-батько VP це S, тоді позначаємо головну складову вузла VP головною складовою вузла S. Якщо вузол-батько VP – це FULLS, FULLQ або SHORTS, тоді переходимо до *кроку 3*, інакше до *кроку 4*.

Крок 3. Позначаємо головну складову вузла VP головною складовою вузла FULLS, FULLQ або SHORTS.

Крок 4. Шукаємо вузол-батько вузла S. Якщо вузол-батько S це FULLS, FULLQ або SHORTS, тоді позначаємо головну складову вузла S головною складовою вузла FULLS, FULLQ або SHORTS.

Крок 5. Шукаємо у дереві розбору вузол noun (pronoun). Якщо вузол-батько вузла noun (pronoun) є NP, тоді позначаємо noun (pronoun) головною складовою вузла NP, а сам вузол NP позначаємо SUBJECT. Переходимо до *кроку 6*.

Крок 6. Шукаємо у дереві розбору вузол noun (pronoun):

- якщо вузол-батько вузла noun (pronoun) є NG, тоді позначаємо noun (pronoun) головною складовою вузла NG, а сам вузол NG позначаємо OBJECT;
- якщо вузол-батько вузла noun (pronoun) є DNP, тоді позначаємо noun (pronoun) головною складовою вузла DNP. Якщо вузол-батько вузла DNP є TARGET, тоді позначаємо DNP головною складовою вузла TARGET, а сам вузол TARGET позначаємо OBJECT;
- якщо вузол-батько вузла noun (pronoun) є ADDRESS, тоді позначаємо noun (pronoun) головною складовою вузла ADDRESS, а сам вузол ADDRESS позначаємо OBJECT.

Крок 7. Якщо не всі вершини дерева позначені, тоді перейти до кроку 5, інакше кінець алгоритму.

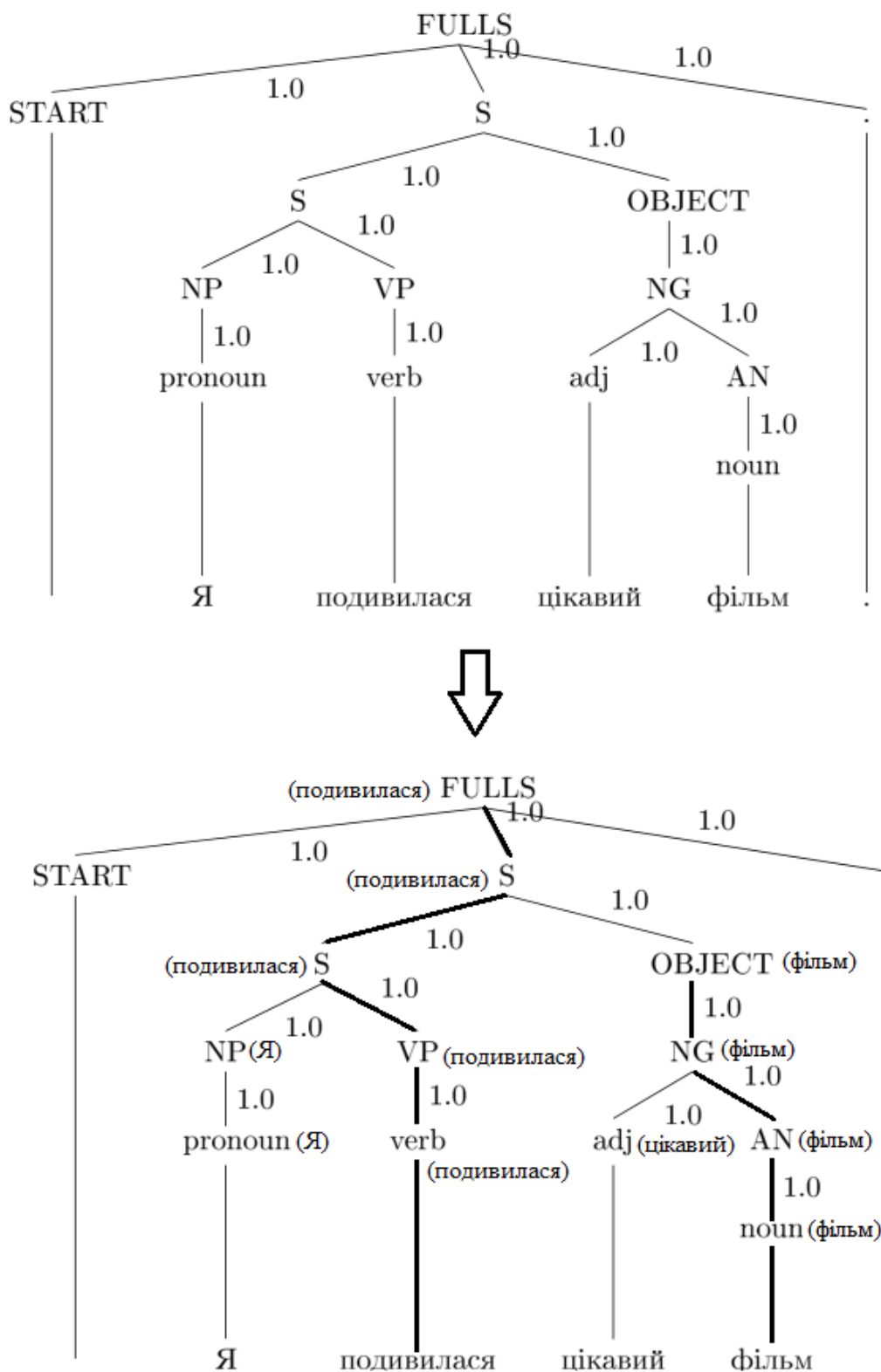


Рис. 3.19. Визначення головної складової у дереві складових для речення УСМ «Я подивилася цікавий фільм»

Наведений алгоритм дає змогу перетворити дерева семантично-синтаксичного розбору, які отримані за допомогою парсеру APCFG, на дерева залежностей. Після цього можна перекладати речення української словесної мови на українську жестову мову та навпаки, використовуючи основні правила граматики (наведені у розділі 3.2).

Так, можна побудувати дерева граматики залежностей. Згідно з описаним алгоритмом трансформацію дерева складових у дерево залежностей для речення УСМ «Я подивилася цікавий фільм» подано на рис. 3.20.

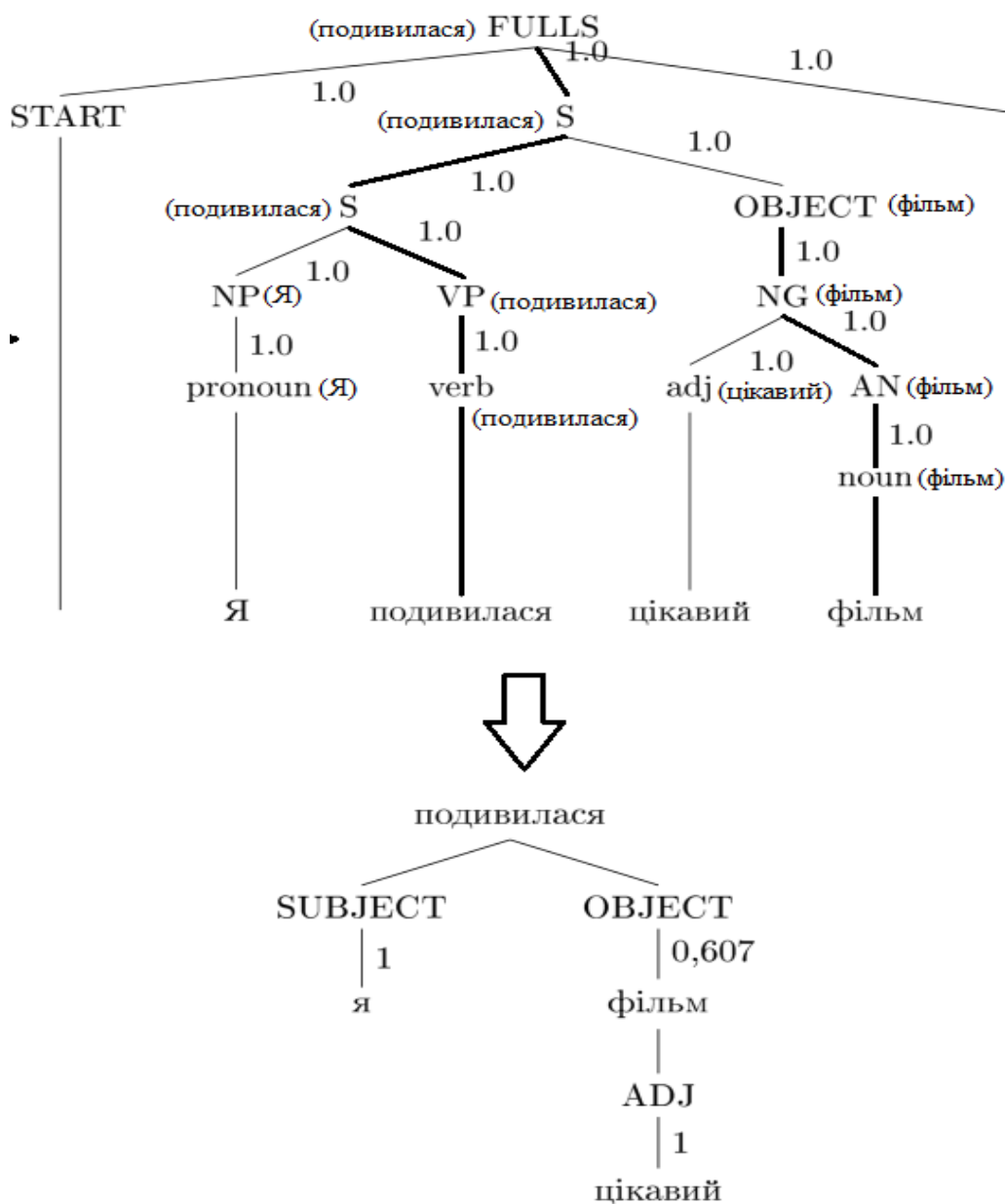


Рис. 3.20. Трансформація дерева складових у дерево залежностей для речення УСМ «Я подивилася цікавий фільм»

Цей алгоритм перетворення дерева складових на дерево залежностей протестований на 220 реченнях УСМ корпусу паралельних текстів «Українська словесна мова – українська жестова мова». Результати тестування алгоритму трансформації дерева складових у дерево залежностей показали високу ефективність застосування цього алгоритму (92 % правильно перетворених речень).

Для питальних речень української словесної мови алгоритм визначення головної складової такий, як і для розповідних речень. Наприклад, для речення УСМ «Ти ходиш до школи?» визначення головної складової та трансформацію дерева складових у дерево залежностей подано на рис. 3.21.

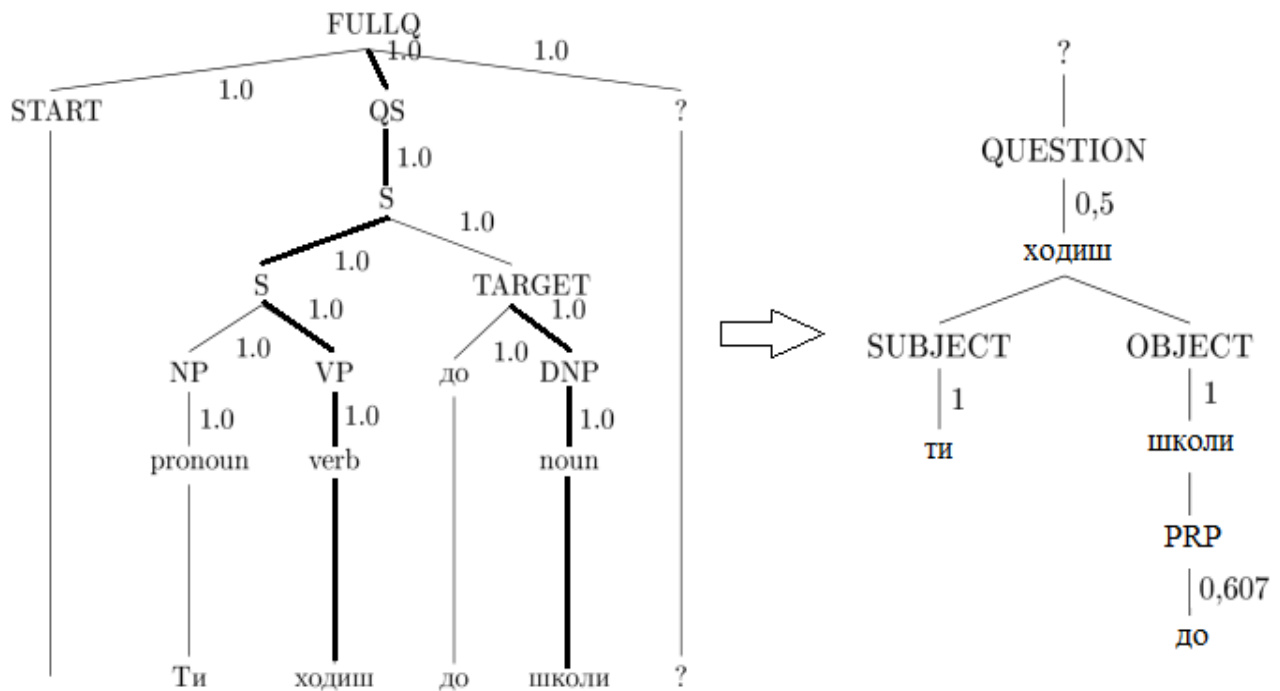


Рис. 3.21. Трансформація дерева складових у дерево залежностей для питального речення УСМ «Ти ходиш до школи?»

Алгоритм визначення головної складової у дереві складових добре працює на тих реченнях УСМ, у яких чітко визначено порядок слів. Проте, є випадки, коли порядок слів у реченні української словесної мови такий, що під час розбору не завжди можна прикріпити залежні слова до головної складової.

Наприклад, у реченні «Гарні прийшли студенти на пари» слово «гарні» є виділене окремим вузлом і є незалежним. Як видно з рис. 3.22, згідно з наведеним алгоритмом, слово «гарні» стає залежним від вузла «прийшли», що є некоректним (слово «гарні» залежне від вузла «студенти»).

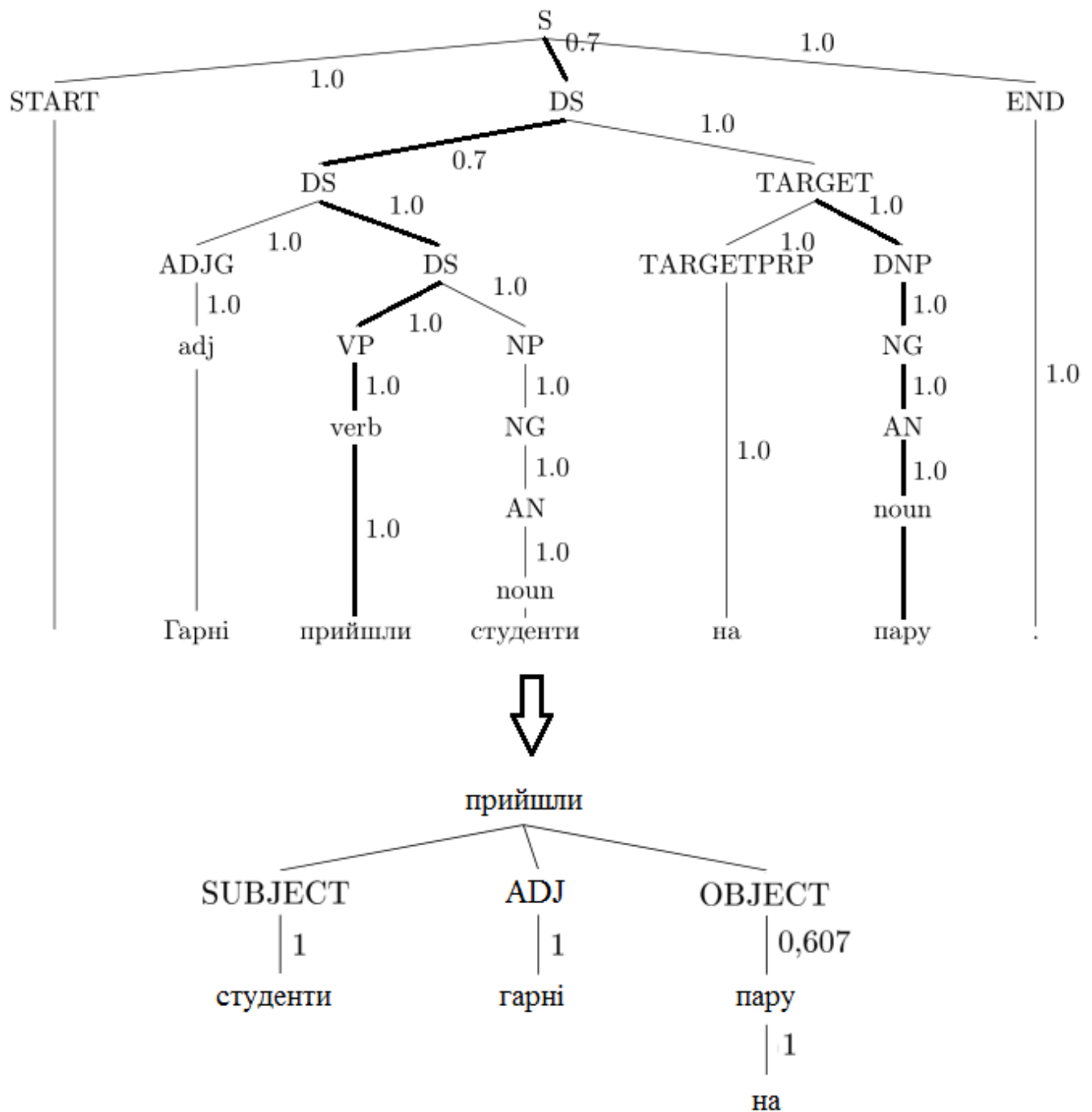


Рис. 3.22. Некоректна трансформація дерева складових у дерево залежностей для речення УСМ «Гарні прийшли студенти на пару»

У такому разі алгоритм трансформації дерева складових у дерево залежностей дає некоректний результат. Правильне дерево залежностей подано на рис. 3.23.

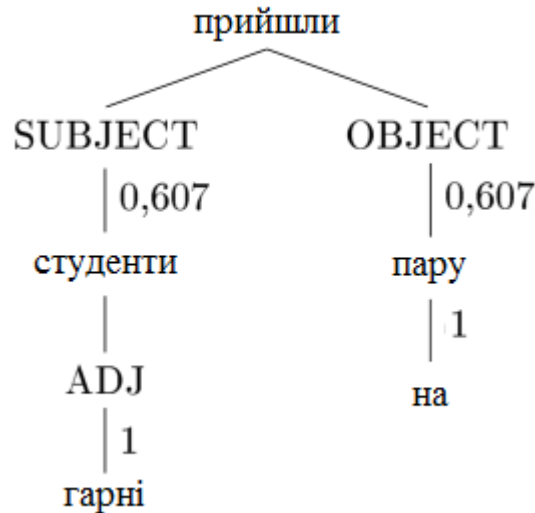


Рис. 3.23. Дерево залежностей для речення УСМ
«Гарні прийшли студенти на пару»

Висновки до третього розділу

У третьому розділі застосовано правила обходу дерев для синтезу речень українською жестовою мовою. Описано основні правила перекладу анотованої української жестової мови на словесну та навпаки. Для семантично-синтаксичного розбору речень української словесної та анотованої української жестової мов використано метод розбору на основі граматики APCFG. Розроблено та систематизовано основні групи правил для семантично-синтаксичного розбору речень української словесної мови та анотованої УЖМ. Для речень УСМ, розроблені правила семантично-синтаксичного розбору, поділено на 5 груп (група правил для побудови речень, група правил для іменникових синсетів, група правил для дієслівних синсетів, група правил для питальних речень, група правил для речень, у яких порушено порядок слів). Для семантично-синтаксичного розбору речень анотованої УЖМ правила

поділяються на чотири групи (відсутня п'ята група правил, у яких порушено порядок слів, оскільки для УЖМ порядок слів у реченні чітко визначений) та аналогічні до перших чотирьох груп української словесної мови. Розроблений дисертантом парсер на основі афіксної ймовірнісної контекстно-вільної граматики дає змогу здійснити семантичний аналіз речень української словесної та анотованої української жестової мов.

Для перекладу речень української словесної мови на речення української жестової мови та навпаки використано метод трансформації дерева складових (семантично-синтаксичного розбору) у дерево залежностей.

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, які розробила дисертант.

Практична апробація та аналіз одержаних результатів підтвердили доцільність використання наведених методів для перекладу анотованої української жестової мови.

РОЗДІЛ 4

ПРОГРАМНО-АЛГОРИТМІЧНИЙ КОМПЛЕКС ПЕРЕКЛАДУ АНОТОВАНОЇ УКРАЇНСЬКОЇ ЖЕСТОВОЇ МОВИ

У четвертому розділі подано схему роботи програмно-алгоритмічного комплексу перекладу анотованої української жестової мови з використанням граматично доповненої онтології. Спроековано архітектуру програмно-алгоритмічного комплексу, який містить такі функціональні програмно-алгоритмічні модулі: модуль наповнення граматично доповненої онтології, модуль перекладу із застосуванням граматично доповненої онтології, модуль тестування перекладу. Описано структуру програмно-алгоритмічних модулів та їх призначення. Розроблено структури баз даних (БД граматично доповненої онтології, БД правил граматики APCFG, БД правил перекладу, БД тестувальних речень), які містить ПАК. Наведено алгоритм перекладу анотованої української жестової мови з використанням ГДО. Проведено експерименти та наведено результати тестування розробленого ПАК двостороннього перекладу анотованої української жестової мови.

4.1. Схема роботи програмно-алгоритмічного комплексу перекладу анотованої української жестової мови

Для розроблення програмно-алгоритмічного комплексу перекладу анотованої української жестової мови на основі граматично доповненої онтології використано сучасні підходи до створення комп'ютерних систем перекладу жестових мов. У межах дисертаційного дослідження автор спроектувала структуру програмно-алгоритмічного комплексу, побудувала систему правил для перекладу анотованої української жестової мови з використанням граматично доповненої онтології, розробила алгоритми на основі досліджуваних методів перекладу української жестової мови, провела експериментальні дослідження та виконала порівняльний аналіз отриманих результатів.

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

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

Основною метою функціонування програмно-алгоритмічного комплексу для перекладу анотованої української жестової мови є двосторонній переклад тексту, а саме – введення тексту, його розбиття на речення, переклад речень та їх виведення у вигляді тексту як відеозапису жестикулювання (якщо переклад здійснюється на українську жестову мову, то виведення жестів з використанням відео).

Призначенням розробленого програмно-алгоритмічного комплексу є максимальна автоматизація перекладу української жестової мови на українську словесну та навпаки. Інтерфейс програмно-алгоритмічного комплексу є зручним та доступним як для чуючих, так і для нечуючих користувачів з навичками

роботи на комп'ютері, та дає змогу комфортно реалізувати двосторонній переклад УЖМ.

Для здійснення перекладу за допомогою програмно-алгоритмічного комплексу нечуючий користувач використовує клавіатуру для введення речення, яке потрібно перекласти, монітор, мишку для вибору перекладу тощо. Для чуючого користувача передбачене використання таких технічних засобів для роботи з ПАК як монітор, клавіатура, мишка.

Схема роботи ПАК перекладу з використанням граматично доповненої онтології зображена на рис. 4.1.

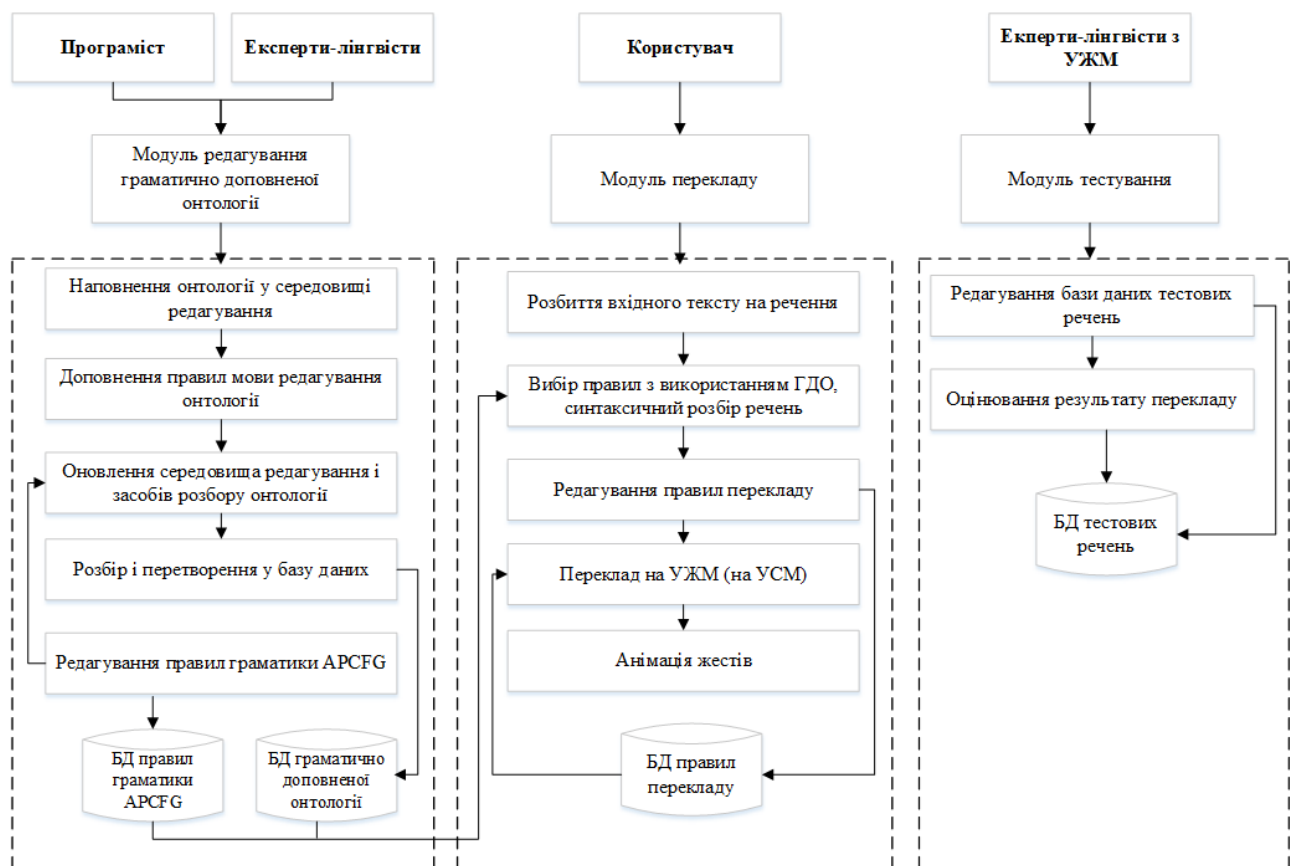


Рис. 4.1. Схема роботи ПАК перекладу на УЖМ з використанням граматично доповненої онтології

Процеси, що реалізовані в інформаційній технології перекладу, подані діаграмою варіантів використання на рис. 4.2.

Процес наповнення граматично доповненої онтології відбувається із залученням експертів-лінгвістів, програмістів та експертів з української жестової мови.

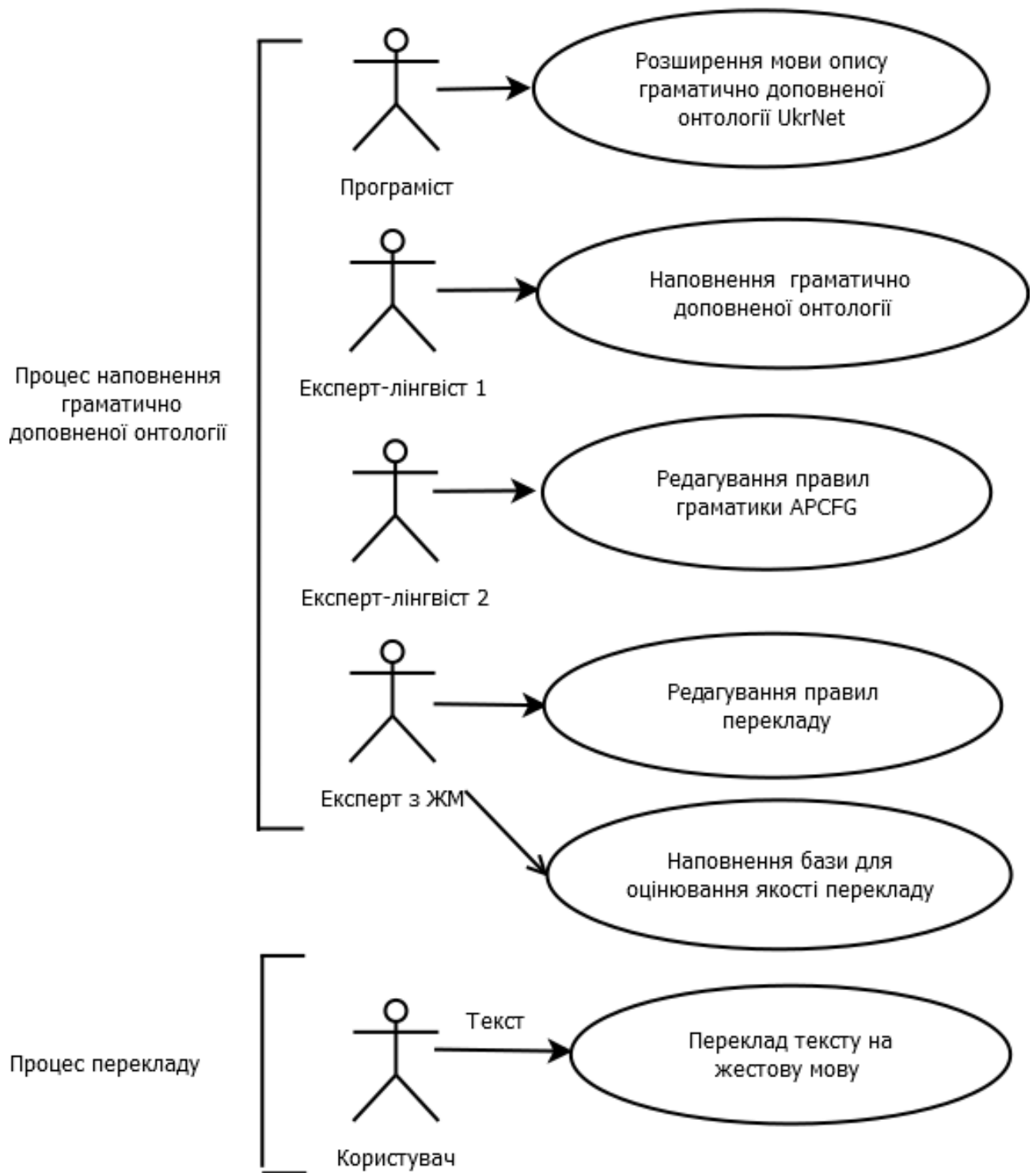


Рис. 4.2. Діаграма варіантів використання інформаційної технології перекладу

Розширення мови опису граматично доповненої онтології та можливості введення нових понять здійснює програміст. Наповнення граматично доповненої онтології (вибір тематики, слів, жестів, побудова граматичних конструкцій) здійснюється із залученням експертів-лінгвістів (одного або декількох). Для редагування правил граматики APCFG, за допомогою яких здійснюється семантично-синтаксичний розбір речень, залучається експерт-лінгвіст.

Експерт з української жестової мови задіяний для редагування правил двостороннього перекладу УЖМ, а також для наповнення бази тестувальних речень для оцінювання якості перекладу.

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

Модуль перекладу складається із програмних пакетів, призначених для вирішення завдань введення, виведення та опрацювання даних на комп'ютері, а саме:

- пакет семантично-синтаксичного розбору *UkrParser* речень української словесної мови та української жестової мови (com.langproc);
- пакет словників української словесної мови та української жестової мови (org.dts.spell);
- пакет синтезу правил з онтології для семантично-синтаксичного розбору речень (org.ontologytoparse);
- пакет синтезу української жестової мови з використанням відеословника (org.signvideo).

Модуль наповнення граматично доповненої онтології складається із програмного пакета org.gaodl, який дає змогу редагувати мову опису ГДО та введення нових понять.

Модуль тестування перекладу складається із програмних пакета оцінювання результату перекладу (org.estimationetnslate).

4.1.1. Інструкція користувачеві

Переклад анотованої української жестової мови здійснює користувач, який вибирає мову введення речень та мову, на яку будуть перекладатись введені ним речення (рис. 4.3). Після цього користувач натискає кнопку «Переклад», яка викликає процедуру семантично-синтаксичного розбору введеного речення, процедуру трансформації дерева складових у дерево залежностей, процедуру синтезу речень з використанням правил граматики та процедуру виведення речень.

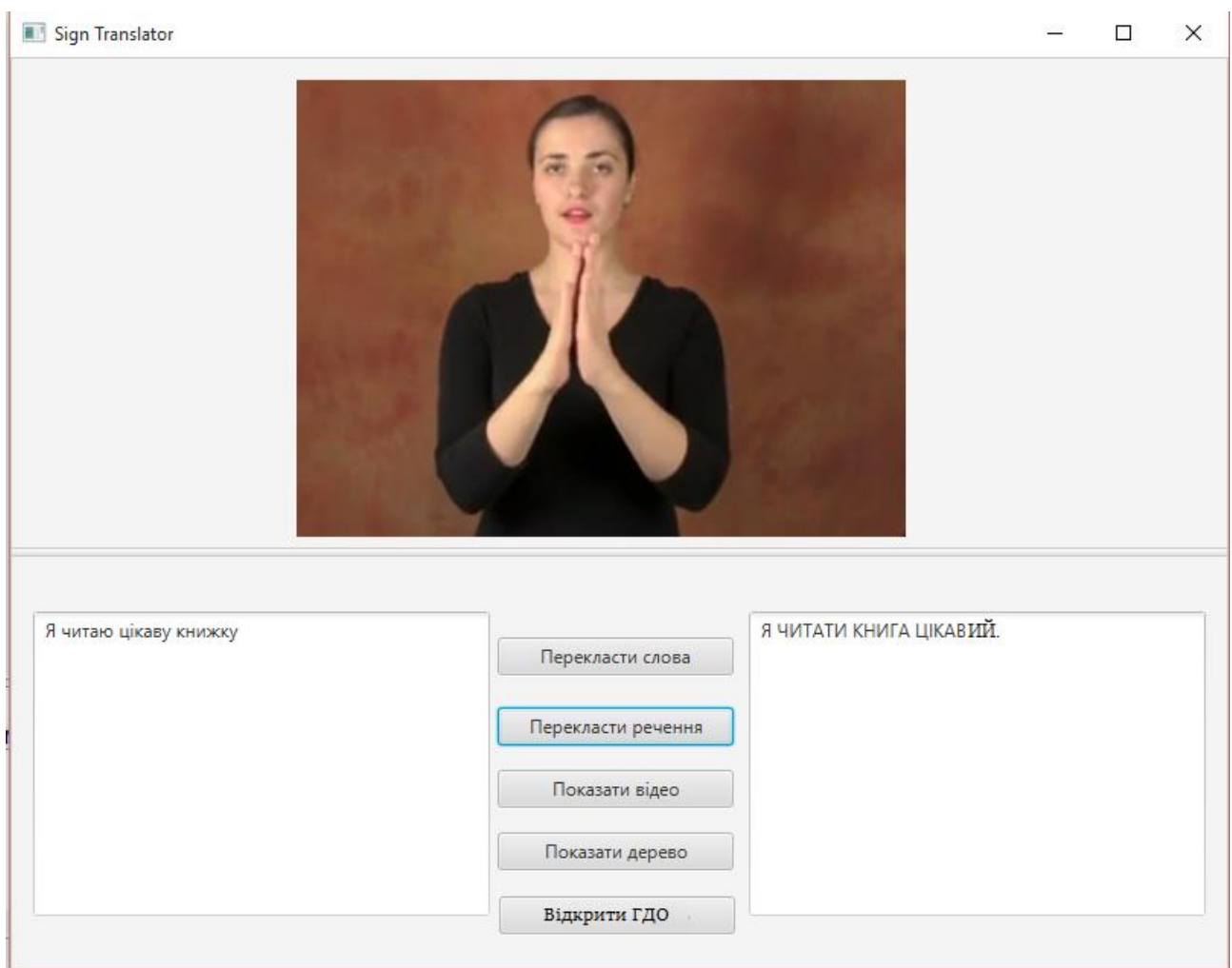


Рис. 4.3. Інтерфейс ПАК перекладу анотованої української жестової мови

Крім того, користувач має змогу переглянути процедуру семантично-синтаксичного аналізу речень, натиснувши кнопку «Показати дерево».

Семантично-синтаксичний аналіз речень подано у вигляді дерева складових, яке виводиться у програмі (рис. 4.4).

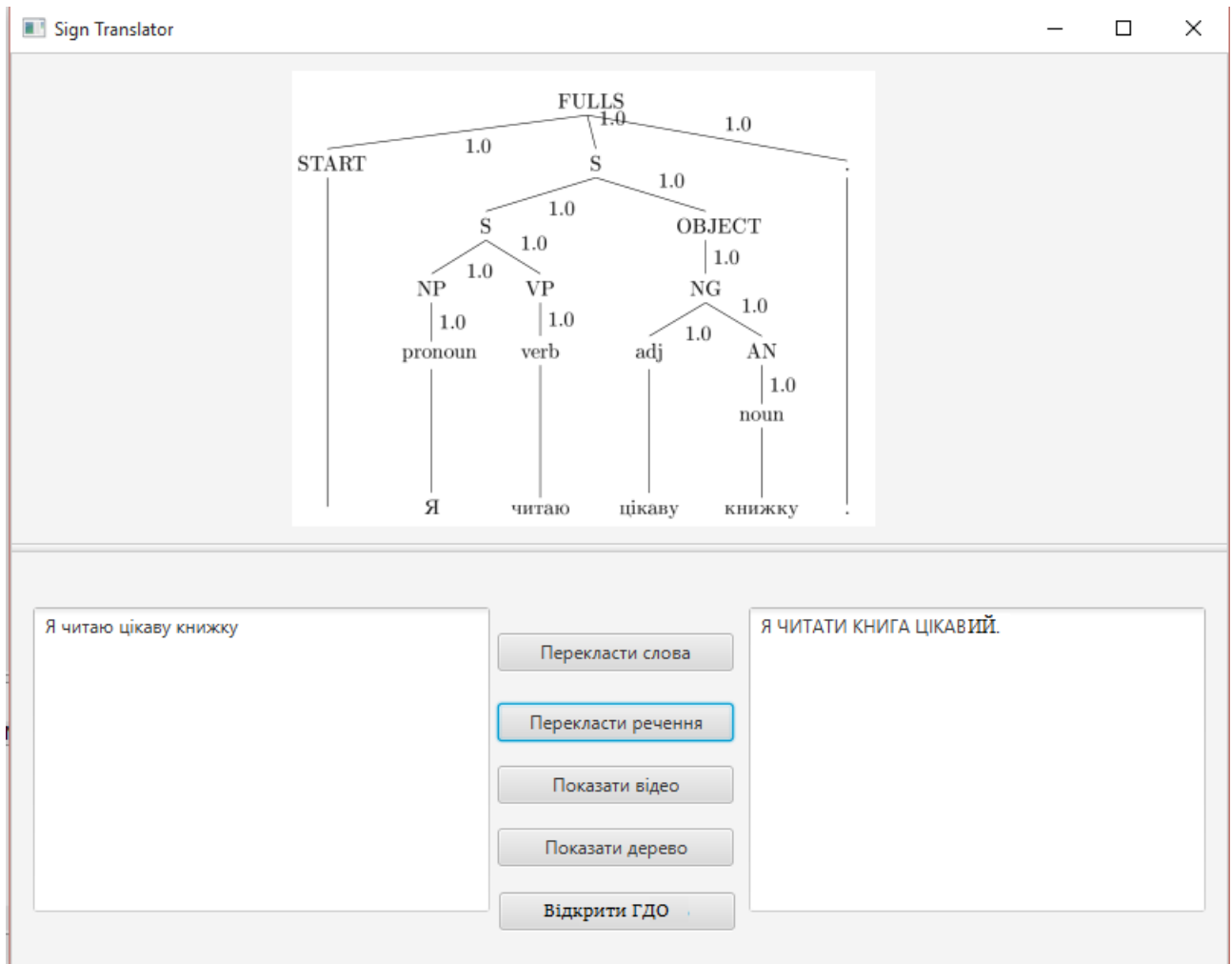


Рис. 4.4. Дерево семантично-синтаксичного розбору речення «Я читаю цікаву книжку»

Натиснувши кнопку «Показати відео», користувач має змогу переглянути речення українською жестовою мовою у вигляді відеозапису жестикулювання (рис. 4. 2).

Програміст має змогу доповнити онтологію новими поняттями, натиснувши кнопку «Відкрити ГДО». Наповнення граматично доповненої онтології відбувається в середовищі Eclipse. Для цього потрібно завантажити модуль редагування ГДО і встановити додаток Xtext (рис. 4.5).

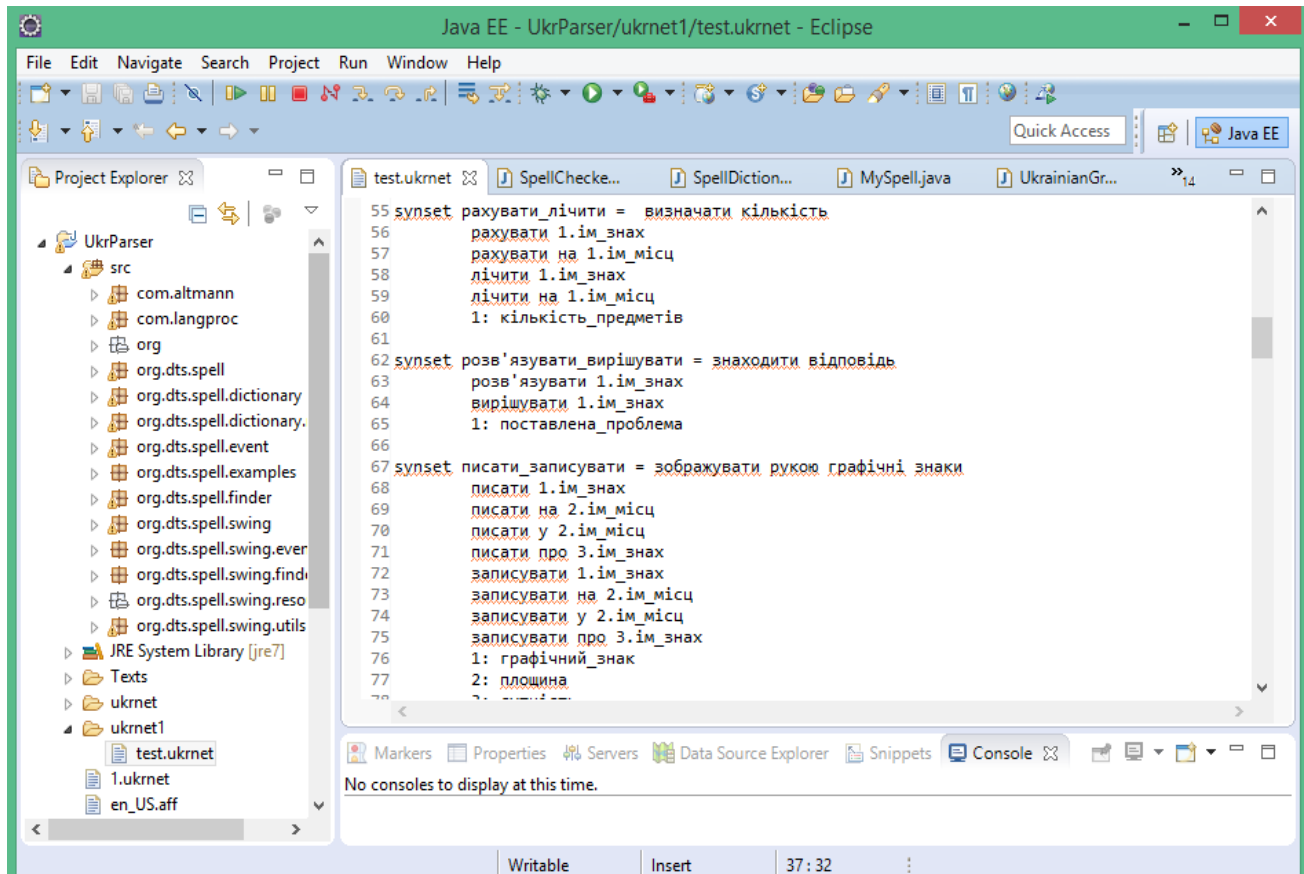


Рис. 4.5. Файл для редагування ГДО

Програмно-алгоритмічний комплекс перекладу анотованої української жестової мови здійснює переклад за таким алгоритмом:

- 1) для початку роботи програмно-алгоритмічного комплексу користувач визначає, з якої мови буде здійснюватись переклад;
- 2) якщо з української жестової мови – вказується яким способом буде вводиться інформація: кожне слово речення УЖМ вводиться у нотації глосів за допомогою клавіатури;
- 3) переклад із використання граматично доповненої онтології. Програмно-алгоритмічний комплекс містить модуль перекладу, який використовує граматично доповнену онтологію для семантично-синтаксичного розбору речень та їх перекладу. Також розроблено модуль для наповнення ГДО та редагування правил граматики APCFG;
- 4) вивід перекладеного тексту. Якщо переклад здійснювався з української словесної мови на анотовану українську жестову мову, то

забезпечується показ жестів з бази відеозаписів, а також є можливість під'єднання модуля синтезу жестової мови з використанням аватара.

Блок-схема алгоритму перекладу на українську жестову мову зображена на рис. 4.6.

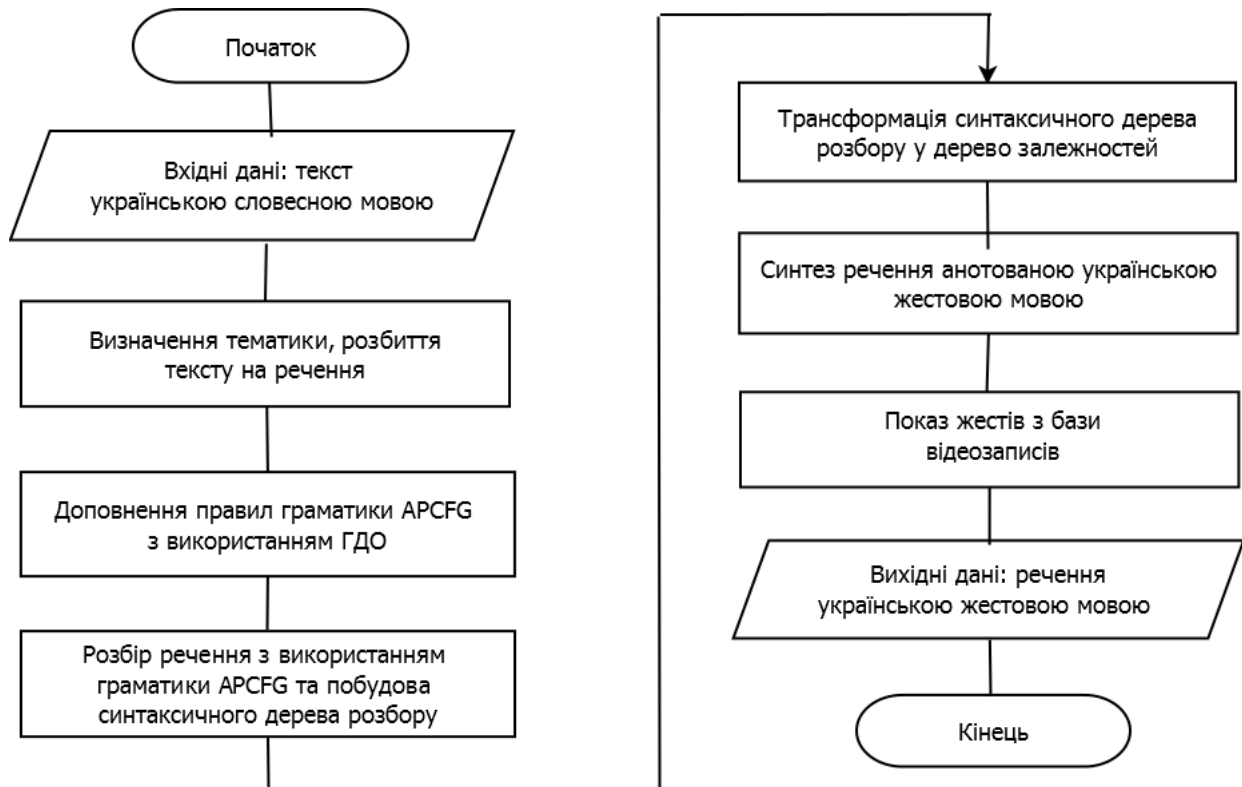


Рис. 4.6. Блок-схема алгоритму перекладу українською жестовою мовою

Розглянемо роботу ПАК, коли переклад здійснюється з української словесної мови на українську жестову. На вхід ПАК подається текст українською словесною мовою. Після введення тексту, відбувається його розбиття на речення та визначення їхньої тематики («Навчання», «Природа», «Подорож», «Держава», «Сім'я», «Виробництво», «Професії», «Військо», «Театр», «Культура», «Лікарня»). На наступному етапі здійснюється пошук правил у граматично доповненій онтології, за якими можна побудувати дерево семантично-синтаксичного розбору для цього речення. Цей етап полягає у розборі речення з використанням афіксної ймовірнісної контекстно-вільної граматики та здійснюється із використанням бази граматично доповненої онтології та бази правил граматики APCFG. У результаті цього будується дерево

семантично-синтаксичного розбору речення. Після семантично-синтаксичного розбору речення виконується переклад цього речення на УЖМ. Переклад здійснюється із залученням бази правил перекладу, яка складається з правил перекладу на жестову мову, правил порядку слів у реченнях відповідно до УЖМ.

Оцінюють результат перекладу через порівняння речень з наявними перекладами у базі тестових речень. Якщо цього речення немає у базі, то здійснюється його додавання у базу та відповідна перевірка правильності перекладу із залученням експерта-лінгвіста з УЖМ.

База правил граматики APCFG та база граматично доповненої онтології постійно оновлюються через додавання нових правил та нових синсетів відповідно.

4.2. Структура баз даних програмно-алгоритмічного комплексу

Схема роботи програмно-алгоритмічного комплексу для перекладу української жестової мови, зображена на рис. 4.2, передбачає наявність чотирьох баз даних: БД граматично доповненої онтології, БД правил граматики APCFG, БД правил перекладу, БД тестувальних речень.

4.2.1. База даних граматично доповненої онтології

Структура БД граматично доповненої онтології містить такі сутності (рис. 4.7): «Параметризоване висловлювання», «Предикат», «Слоти предиката», «Додатки предиката», «Граматична форма додатка», «Відношення між предикатами», «Відношення між додатками».

Сутність «Параметризоване висловлювання» в цій схемі БД відповідає певному висловлюванню (предмета, дії, ознаці тощо), яке має подання в українській словесній мові та в українській жестовій мові. Ця сутність має такі атрибути: «№ Парам_висл» (первинний ключ) та «Парам_висл» параметризованого висловлювання.

Параметризоване висловлювання в українській словесній мові виражається певним словосполученням, яке містить предикат та додатки предиката. Сутність

«Предикат» має такі атрибути: «№ Предиката» (первинний ключ), «Предикат» та зовнішні ключі «Відн_предикатами» і «Парам_висл».

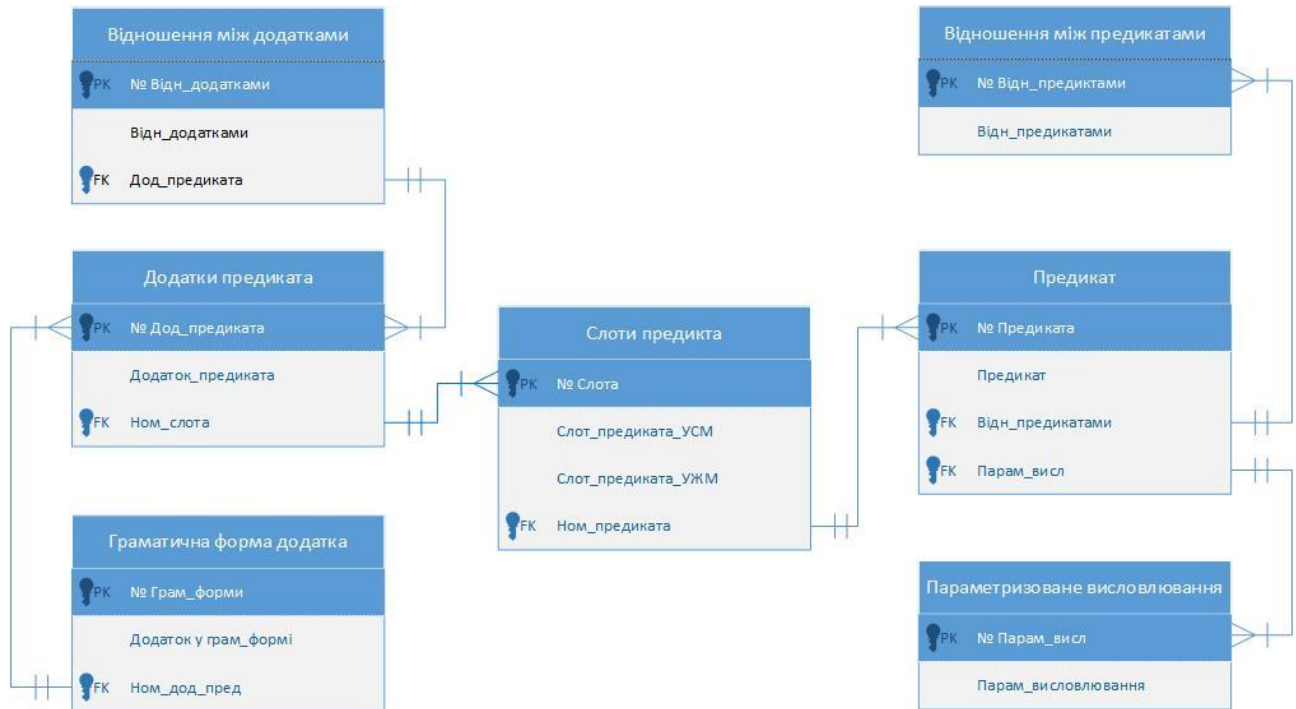


Рис. 4.7. Діаграма «сутність-зв'язок» БД граматично доповненої онтології української словесної та жестової мов

Наприклад, перелік значень для сутності Предикат (№ Предиката, Предикат, Відн_предикатами, Парам_висл) = {(1, учити_навчати, 3, 4), (2, учити_вивчати, 1, 5), (3, вчитися_у, 4, 7)}.

Сутність «Слоти предиката» містить атрибути «№ Слота», який є первинним ключем, «Слот_предиката_УСМ», «Слот_предиката_УЖМ» та зовнішній ключ «Ном_предиката». Слоти предиката містять граматичні конструкції вираження предиката як в українській словесній мові, так і в українській жестовій мові. Як приклад, наведено перелік значень для сутності Слоти предиката (№ Слота, Слот_предиката_УСМ, Слот_предиката_УЖМ, Ном_предиката) = {(1, давати_поради, жест_навчати, 1), (2, опанувати, жест_вивчати, 2), (3, відвідувати, жест_вчитися, 3)}.

Сутність «Додатки предиката» містить ключовий атрибут «№ Дод_пред», «Додаток_предиката» та зовнішній ключ «Ном_слота». Згідно з математичною

моделлю граматично доповненої онтології, додатки предиката мають свій чітко визначений порядковий номер у граматичній конструкції параметризованого висловлювання. Додатками предиката є звичайні слова, і кожне слово вказано в певній граматичній формі. Наприклад, перелік значень для сутності Додатки предиката (№ Дод_предиката, Додаток_предиката, Ном_слота) = {(1, 1.ім_знах, 1), (1, 2.ім_дав, 1), (1, 1.ім_жест, 1), (1, 1.ім_жест, 1), (2, 1.ім_знах, 2), (2, 1.ім_жест, 2), (3, 1.ім_місц, 3), (3, 1.ім_знах, 3), (3, 1.ім_жест, 3)}.

Сутність «Грамматична форма додатка» містить ключовий атрибут «№ Грам_форми», атрибути «Додаток у граматичній формі» та «Ном_дод_пред».

Сутність «Відношення між предикатами» має такі атрибути «№ Відн_пред» (первинний ключ) та «Відн_пред». Серед відношень на предикатах розглянуто такі: тропонімії, антонімії, імплікації, гіперонімії/гопонімії та ін. Наприклад, перелік значень для сутності Відношення між предикатами (№ Відн_пред, Відношення_пред) = {(1, common), (2, optional_parts), (3, associations)}.

Сутність «Відношення між додатками» містить ключовий атрибут «№ Відн_дод» та атрибути «Відн_дод» і «Дод_предиката». Між додатками реалізовано відношення гіперонімії/гопонімії, меронімії/голонімії, антонімії, асоціації та ін. Як приклад, наведено перелік значень для сутності Відношення між додатками (№ Відн_дод, Відн_дод, Дод_предиката) = {(1, common, 1), (2, optional_parts, 2), (3, associations, 1), (4, optional_associations, 2)}.

4.2.2. База даних правил перекладу

Структура БД правил перекладу містить такі сутності (рис. 4.8): «Правила перекладу», «Предикат», «Суб'єкт», «Об'єкт», «Означення», «Обставина», «Заперечення», «Питальне слово».

Сутність «Правила перекладу» має такі атрибути: «№ Прав_перекладу» (первинний ключ) та «Тлумачення_правила». Наприклад, перелік значень для сутності Правила перекладу (№ Прав_перекладу, Тлумачення_перекладу) = {(1,

VERB->show (Subject, * , Object)), (2, SUBJECT->show (Adj1, Adj2,..., AdjN, *), (3, OBJECT->show (Adj1, Adj2,..., AdjN, *)}).

Сутність «Предикат» має такі атрибути: «№ Предиката» (первинний ключ), «Предикат» та «Прав_перек». Наведено приклад переліку значень для сутності Предикат (№ Предиката, Предикат, Прав_перек) = {(1, VERB[FUTURE]->show (Subject, * , Object), 1), (2, VERB[PAST]->show (Subject, * , Object), 1)}.

Сутність «Суб'єкт» містить атрибути «№ Суб'єкта» (первинний ключ), «Суб'єкт» та «Прав_перек».

Сутність «Об'єкт» містить ключовий атрибут «№ Об'єкта» та атрибути «Об'єкт» і «Прав_перек».

Сутність «Означення» містить атрибути «№ Означення» (первинний ключ), «Означення» та «Прав_перек».

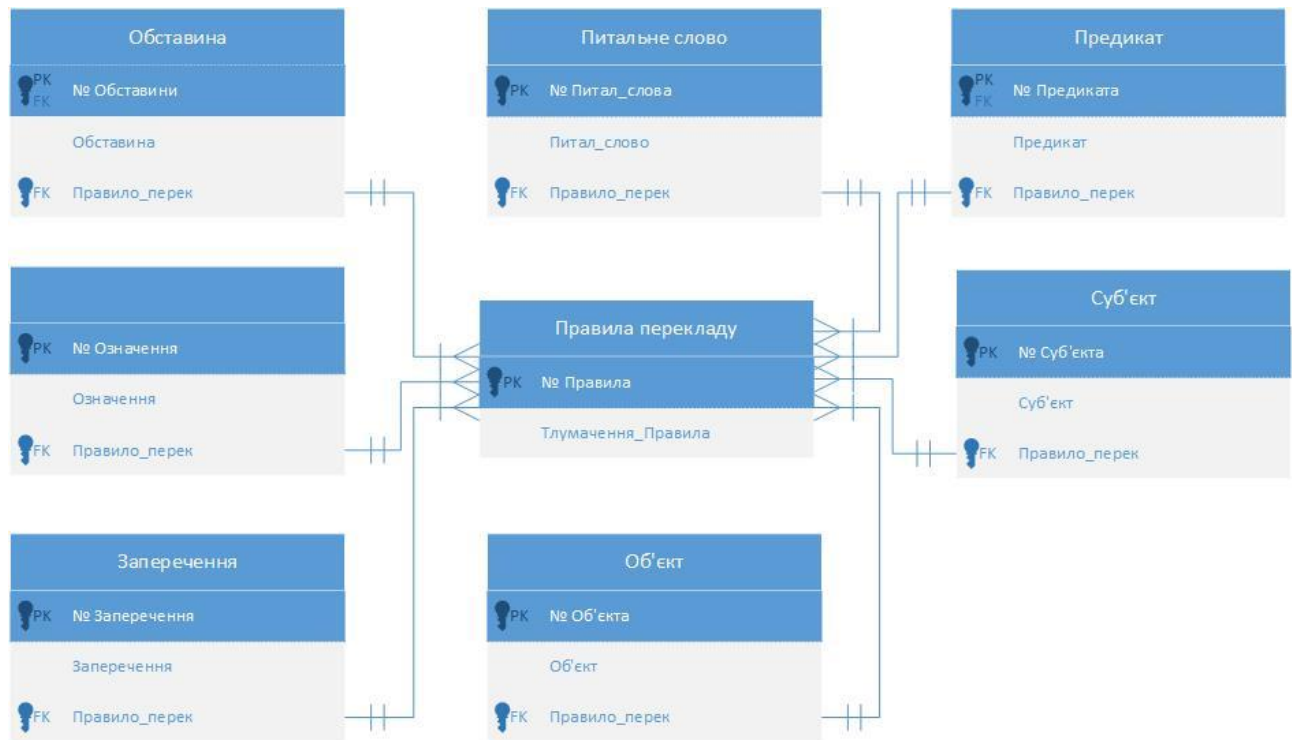


Рис. 4.8. Діаграма «сутність-зв'язок» БД правил перекладу

Сутність «Обставина» містить ключовий атрибут «№ Обставини» та атрибути «Обставина» і «Прав_перек». Наприклад, перелік значень для сутності Обставина (№ Обставини, Обставина, Прав_перек) = {(1, VERB->show (Subject,

ADV, * , Object), 2), (2, VERB[FUTURE]->show (Subject, ADV, * , Object), 2), (3, VERB[PAST]->show (Subject, ADV, * , Object), 2)}.

Сутність «Заперечення» має такі атрибути: «№ Заперечення» (первинний ключ), «Заперечення» та зовнішній ключ «Прав_перек».

Сутність «Питальне слово» містить ключовий атрибут «№ Питал_слова», атрибут «Питал_слово» та зовнішній ключ «Прав_перек». Наприклад, перелік значень для сутності Питальне слово (№ Питал_слова, Питал_слово, Прав_перек) = {(1, VERB[QUESTION]->show (Subject (Pron), * , Object, «?»), 3), (2, VERB[FUTURE, QUESTION]->show (Subject (Pron), * , Object, «?»), 3), (3, VERB[PAST, QUESTION]->show (Subject (Pron), * , Object, «?»), 3)}.

4.2.3. База даних правил граматики APCFG

Структура БД правил граматики APCFG містить такі сутності (рис. 4.9): «Правила граматики APCFG», «Дієслівний синсет», «Іменниковий синсет», «Додаток», «Означення», «Обставина», «Заперечення», «Питальне слово».

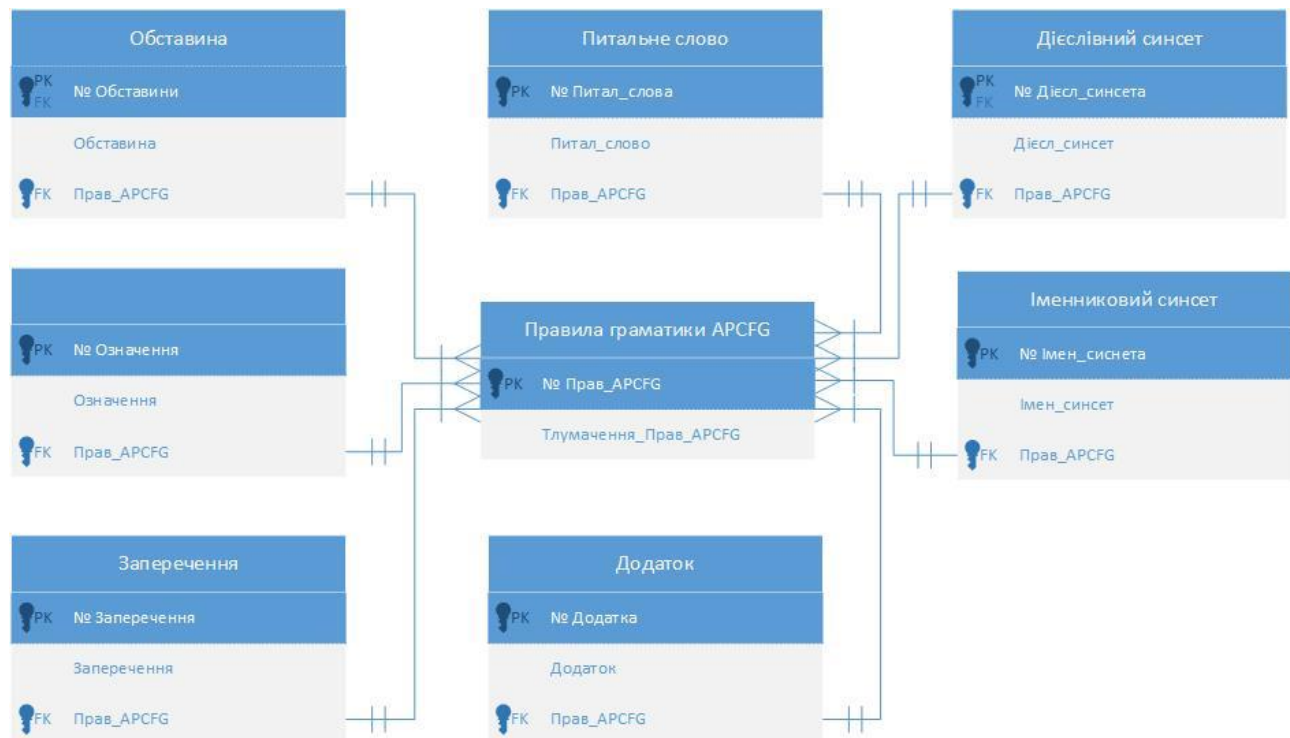


Рис. 4.9. Діаграма «сутність-зв'язок» БД правил граматики APCFG

Сутність «Правила граматики APCFG» має такі атрибути: «№ Прав_ APCFG» (первинний ключ) та «Тлумачення_Прав_ APCFG». Наприклад, перелік значень для сутності Правила граматики APCFG (№ Прав_ APCFG, Тлумачення_Прав_ APCFG) = {(1, FULLS -> START S END), (2, SHORTS -> START NP END)}.

Сутність «Дієслівний синсет» має такі атрибути: «№ Дієсл_синсет» (первинний ключ), «Дієсл_синсет» та зовнішній ключ «Прав_ APCFG». Наприклад, перелік значень для сутності Дієслівний синсет (№ Дієсл_синсет, Дієсл_синсет, Прав_ APCFG) = {(1, verb[PN] → <розповідати-синсет>[PN], 12), (2, verb[PN] → <вчити-синсет>[PN], 56), (3, verb[PN] → <грати-синсет>[PN], 21)}.

Сутність «Іменниковий синсет» містить атрибути «№ Імен_синсета» (первинний ключ), «Імен_синсет» та зовнішній ключ «Прав_ APCFG». Наведено приклад переліку значень для сутності Іменниковий синсет (№ Імен_синсет, Імен_синсет, Прав_ APCFG) = {(1, <людина>[NCG] → <дитина>[NCG], 5), (2, <слухач>[NCG] → <людина>[NCG], 7), (3, <історія-розповідь>[NCG] → <казка>[NCG], 32)}.

Сутність «Додаток» містить ключовий атрибут «№ Додатка» та атрибути «Додаток» і «Прав_ APCFG».

Сутність «Означення» містить атрибути «№ Означення» (первинний ключ), «Означення» та зовнішній ключ «Прав_ APCFG».

Сутність «Обставина» містить ключовий атрибут «№ Обставини» та атрибути «Обставина» і «Прав_ APCFG».

Сутність «Заперечення» має такі атрибути: «№ Заперечення» (первинний ключ), «Заперечення» та зовнішній ключ «Прав_ APCFG». Наприклад, перелік значень для сутності Заперечення (№ Заперечення, Заперечення, Прав_ APCFG) = {(1, VP[PNM] -> <не> VP[PNM], 67), (2, NP[NCG] -> <не> NP[NCG], 73), (3, ADJ[NCG] -> <не> ADJ[NCG], 74)}.

Сутність «Питальне слово» містить ключовий атрибут «№ Питал_слова» та атрибути «Питал_слово» і «Прав_ APCFG».

4.2.4. База даних тестових речень

Структура БД тестових речень українською словесною мовою містить такі сутності (рис. 4.10): «Речення_УСМ», «Словосполучення», «Слово у словосполученні».

Сутність «Речення_УСМ» має такі атрибути: «№ Речення_УСМ» (первинний ключ) та «Речення_УСМ». Наприклад, перелік значень для сутності Речення_УСМ (№ Речення_УСМ, Речення_УСМ) = {(1, Я дуже люблю піцу), (2, Він не любить цукерок)}.

Речення в українській словесній мові виражається одним словом або словосполученням. Сутність «Словосполучення» має такі атрибути: «№ Словосполучення» (первинний ключ), «Словосполучення» та «Речення_УСМ». Перелік значень для сутності Словосполучення (№ Словосполучення, Словосполучення, Речення_УСМ) = {(1, Я люблю, 1), (2, Він не любить, 2)}.

Сутність «Слово у словосполученні» містить атрибути «№ Слова» (порядковий номер слова у словосполученні, є первинним ключем), «№ Словосполучення» та «Слово».

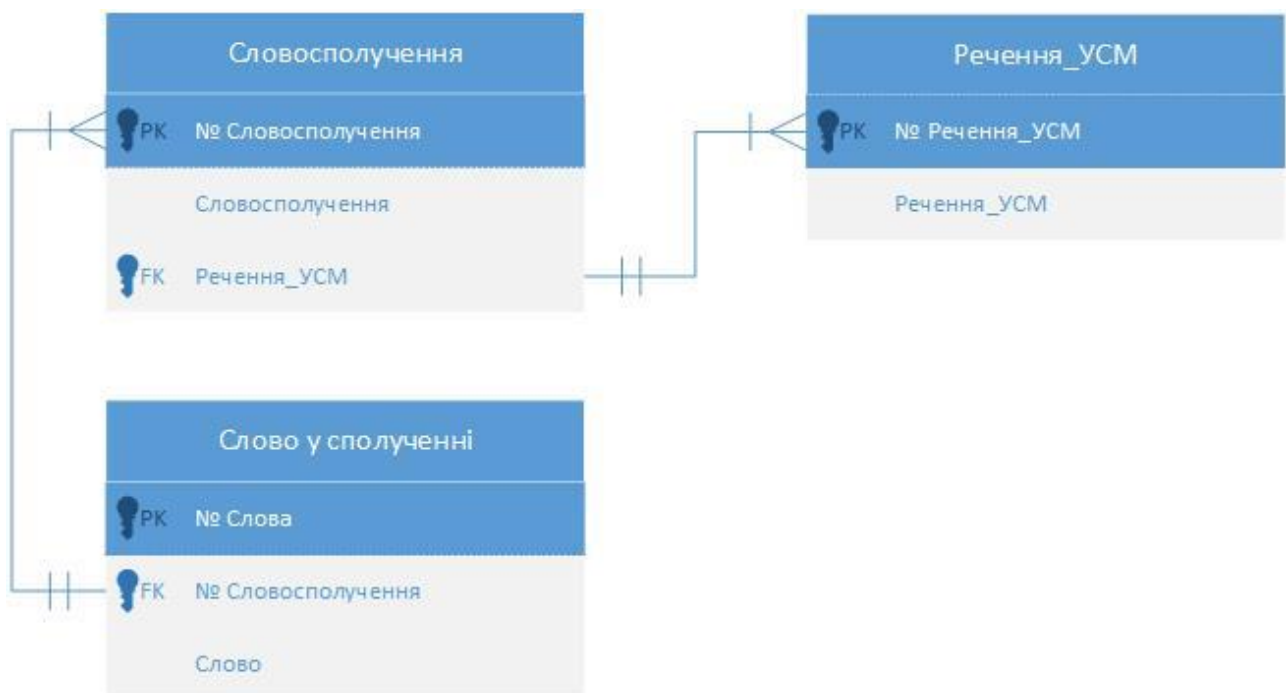


Рис. 4.10. Діаграма «сутність-зв'язок» БД тестувальних речень УСМ

Структура БД тестових речень українською жестовою мовою (рис. 4.11) містить такі сутності: «Речення_УЖМ», «Жестове сполучення», «Жест у сполученні».

Сутність «Речення_УЖМ» має такі атрибути: «№ Речення_УЖМ» (первинний ключ) та «Речення_УЖМ». Наприклад, перелік значень для сутності Речення_УЖМ (№ Речення_УЖМ, Речення_УЖМ) = {(1, Я люблю піца дуже), (2, Він не любить цукерка)}.

У жестовій мові речення виражається одним або декількома синонімічними «Жестовими сполученнями». Жестове сполучення має такі атрибути: «№ Жест_сполучення» (первинний ключ), «Жест_сполучення» та «Речення_УЖМ». Жестове сполучення складається з одного або декількох жестів.

Сутність «Жест у сполученні» містить ключовий атрибут «№ Жесту», «№ Жест_сполучення», «Запис_жесту» (запис жесту у вигляді глоса) та атрибут «Відеозапис».

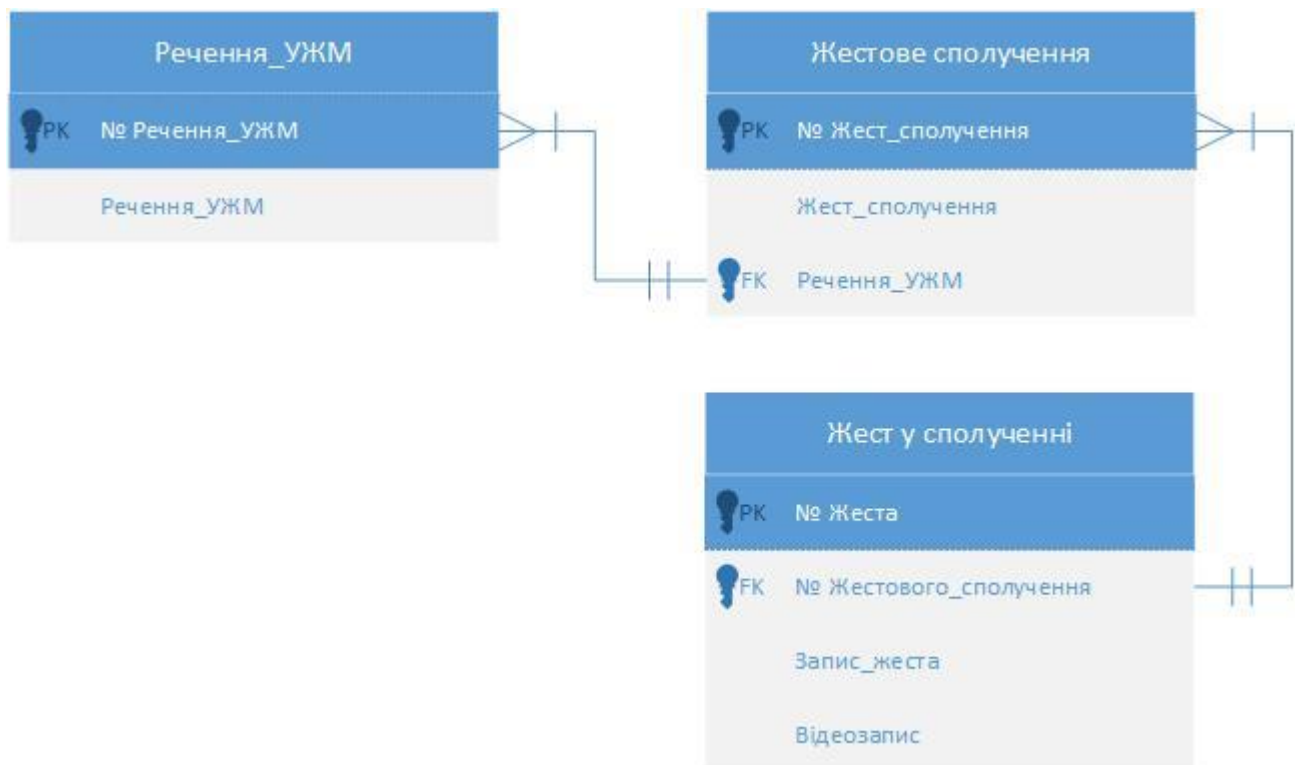


Рис. 4.11. Діаграма «сутність-зв'язок» БД тестових речень УЖМ

4.3. Аналіз результатів тестування системи комп'ютерного перекладу української жестової мови

Для оцінювання якості системи перекладу анотованої української жестової мови проведено чотири експерименти, суть яких полягала в перекладі тестових текстів. Тестові експерименти проводились на 220 реченнях корпусу паралельних текстів «українська словесна мова – українська жести́мова».

У першому експерименті застосовувався лише статистичний методу перекладу української словесної мови на анотовану українську жести́мову та навпаки.

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

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

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

Для аналізу ефективності розроблених методів запропоновано використовувати показник правильно перекладених речень:

$$C = \frac{N_{correct_trans}}{N_{general_trans}} * 100\% ,$$

де $N_{correct_trans}$ – кількість правильно перекладених тестових речень, $N_{general_trans}$ – загальна кількість тестових речень.

Результати оцінювання якості перекладу анотованої української жестової мови наведені у табл. 4.1 та на рис. 4.12.

В експерименті № 1, у якому використовувався статистичний метод машинного перекладу, оцінка якості перекладу речень української словесної

мови у речення анотованої УЖМ становить 57,2 % та 50,5 % під час перекладу речень анотованої УЖМ на речення УСМ.

Таблиця 4.1

Оцінка якості перекладу (%)

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

Експеримент № 2, який було реалізовано з використанням методу машинного перекладу на основі правил, оцінка якості перекладу речень української словесної мови у речення анотованої УЖМ становить 75,3 % та 68,7 % під час перекладу речень анотованої УЖМ на речення УСМ.

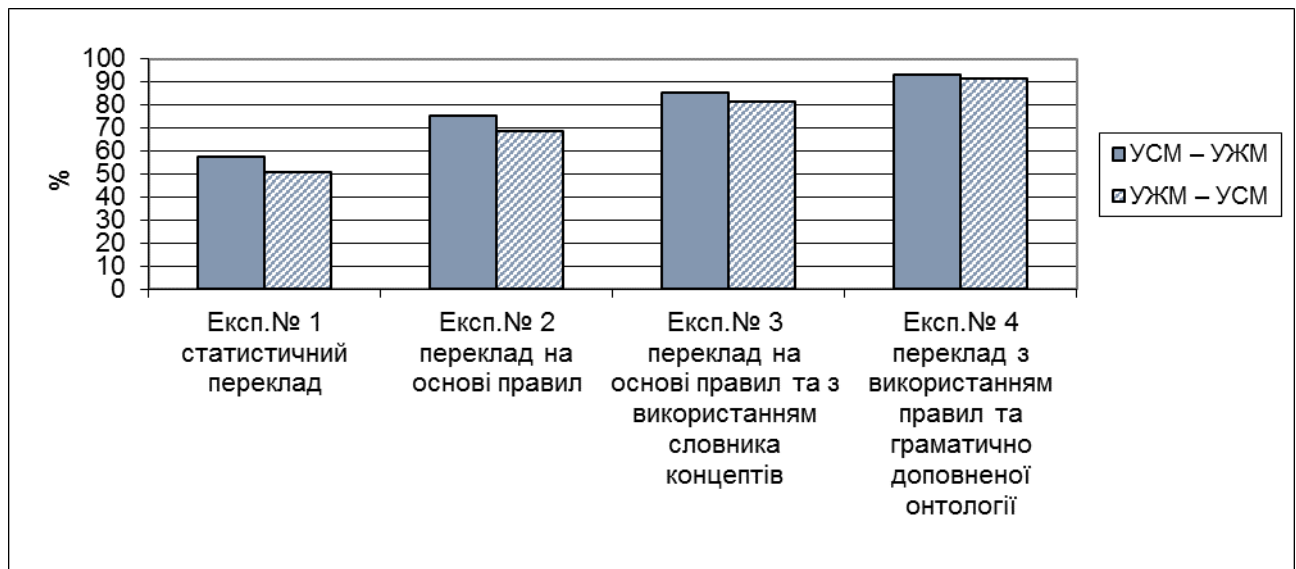


Рис. 4.12. Порівняння методів перекладу української жестової мови

Для експерименту № 3 використано метод машинного перекладу на основі правил з використанням словника концептів, що дало змогу отримати оцінку якості перекладу речень української словесної мови на речення анотованої УЖМ, яка становить 85,5 % та 81,6 % під час перекладу речень анотованої УЖМ на речення УСМ.

В експерименті № 4, який реалізовано з використанням граматично доповненої онтології, оцінка якості перекладу речень української словесної мови у речення анотованої УЖМ становить 93,2 % та 91,4 % під час перекладу речень анотованої УЖМ на речення УСМ.

На основі проведеного оцінювання результату перекладу, помилки перекладу виникають у тих реченнях, для яких семантично-синтаксичний розбір здійснюється неправильно. Наведемо приклади таких речень та дерева семантично-синтаксичного розбору для них (рис. 4.13–4.15).

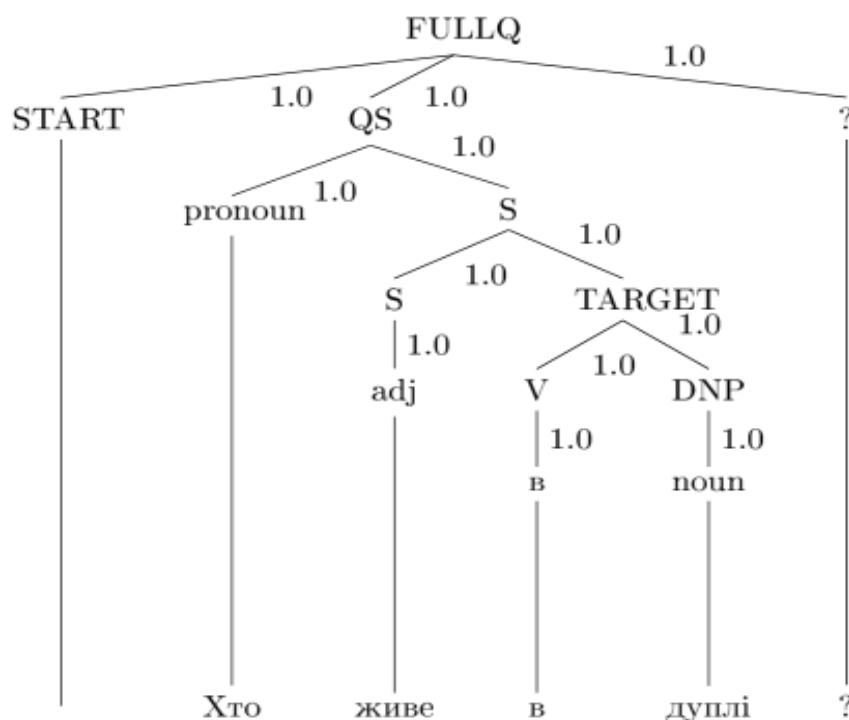


Рис. 4.13. Дерево семантично-синтаксичного розбору речення

«Хто живе в дуплі?»

У реченні «Хто живе в дуплі» було неправильно визначено частину мови для слова «живе» (прикметник замість дієслова).

Для речення «Я була зайнята» слово «була» визначене як прикметник, водночас правильним є варіант дієслова.

У реченні «Наш холодильник білого кольору» частина мови для слова «наш» повинна визначатися як займенник, а у наведеному розборі вказано її як прикметник.

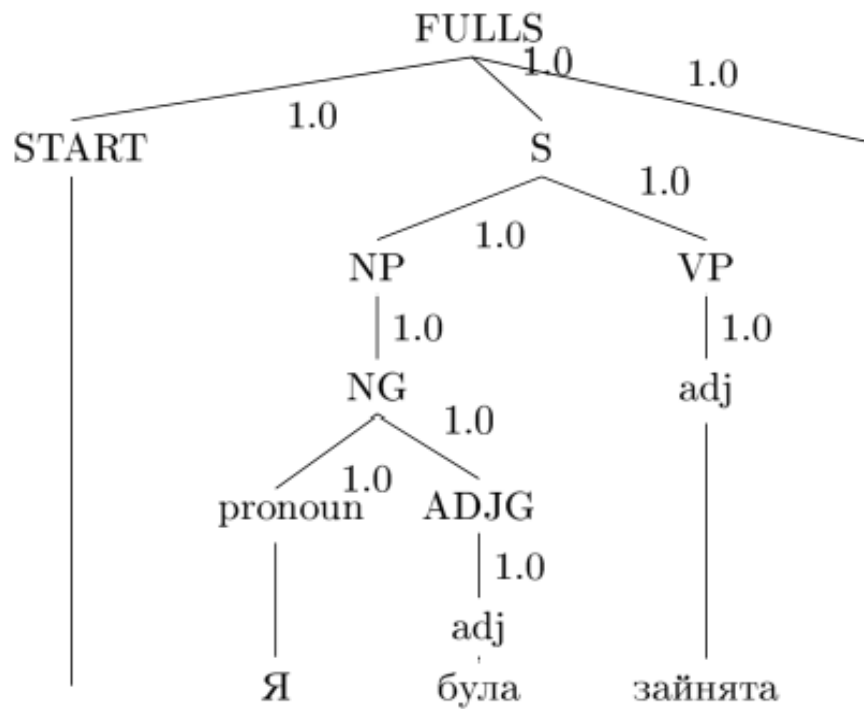


Рис. 4.14. Дерево семантично-синтаксичного розбору речення
«Я була зайнята»

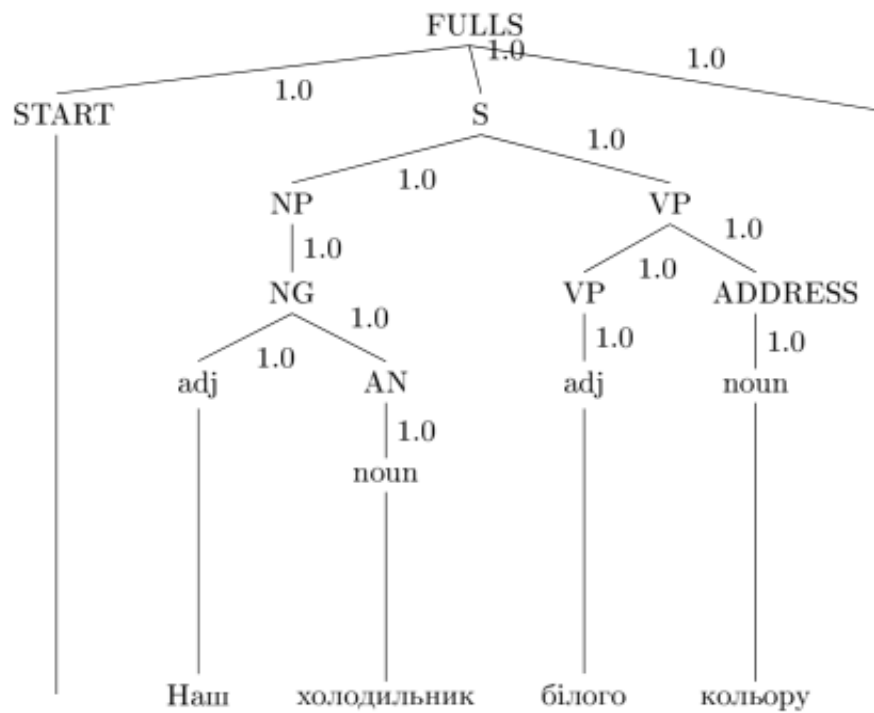


Рис. 4.15. Дерево семантично-синтаксичного розбору речення
«Наш холодильник білого кольору»

Проведене оцінювання результатів роботи ПАК перекладу з використанням граматично доповненої онтології порівняно зі статистичним методом, методом перекладу на основі правил та методом з використанням словника концептів «УСМ-УЖМ» дало змогу покращити якість перекладу до 93,2 %.

Отже, природно зробити висновок, що результат використання розробленої дисертантом системи комп'ютерного перекладу анотованої української жестової мови з використанням граматично доповненої онтології є загалом позитивний. Під час виконання дисертаційного дослідження один із акцентів зроблено на створенні граматично доповненої онтології української словесної та жестової мов для її подальшого використання в системах комп'ютерного перекладу.

Висновки до четвертого розділу

Подальше вдосконалення розроблених технологій перекладу може стати істотним внеском у вирішення проблеми інтеграції людей з вадами слуху в суспільство, зокрема повноцінне навчання нечуючих у вищих навчальних закладах через залучення у студентські групи та шкільні класи.

У результаті дослідження проблеми перекладу анотованої української жестової мови проведено експерименти, які показали, що із застосуванням граматично доповненої онтології розроблено якісну систему комп'ютерного перекладу речень анотованої української жестової мови на речення УСМ (оцінка якості перекладу становить 93,2 %) та навпаки (оцінка якості перекладу становить 91,4 %). У результаті дослідження проблеми розроблено програмно-алгоритмічний комплекс перекладу анотованої української жестової мови з використанням граматично доповненої онтології.

Отримані результати дають можливість застосувати програмно-алгоритмічний комплекс двостороннього перекладу анотованої УЖМ на практиці для навчання осіб з вадами слуху.

Під час дисертаційного дослідження сформовано чотири бази даних (БД граматично доповненої онтології, БД правил граматики APCFG, БД правил перекладу, БД тестових речень), які використовуються під час роботи ПАК.

Програмно-алгоритмічний комплекс реалізовано у вигляді трьох окремих функціональних модулів, зокрема модуля наповнення граматично доповненої онтології, модуля перекладу із застосуванням правил та граматично доповненої онтології та модуля тестування перекладу.

Для подальшого вдосконалення засобів перекладу української жестової мови потрібно розробити модуль розпізнавання речень української жестової мови та вдосконалити методи підготовки даних, доповнити правила семантично-синтаксичного аналізу речень української словесної мови та анотованої УЖМ, враховуючи помилки перекладу.

ВИСНОВКИ

У дисертаційній роботі розв'язано актуальну науково-прикладну задачу 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.

За результатами дослідження, на базі кафедри інформаційних систем та мереж Національного університету «Львівська політехніка» розроблено програмно-алгоритмічний комплекс робочого місця користувача, на якому проведено описані експерименти.

Розроблений автором ПАК використовують для бібліотечного обслуговування нечуючих.

Комп'ютерно-технологічних та методичних аналогів розроблень в Україні поки що немає. Стосовно результатів закордонних досліджень і розроблень подані напрацювання є технологічно співмірними зі світовими досягненнями у галузі новітніх інформаційних технологій, орієнтованих на людей з вадами слуху.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

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. 04, 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». – Х. : 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 // Міжнародний журнал «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 // Міжнар. наук.-техн. журн. «Інформаційні технології та комп'ютерна інженерія». – 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 // *Международ. науч. журн «Science and World»*. – Волгоград, 2014. – № 1(5). – С. 102–105.
 11. 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.
 12. Лозинська О. В. Використання граматично-доповненої онтології для перекладу української жестової мови / О. В. Лозинська // *Зб. ст. «Математика. Інформаційні технології. Освіта», Східноєвропейський національний ун-т ім. Лесі Українки*. – Луцьк, 2015. – № 2. – С. 108–114.
 13. Lozynska O. Using Rule-Based Machine Translation for Ukrainian Sign Language / O. Lozynska, M. Davydov // *Perspective Technologies and Methods in MEMS Design MEMSTECH2014: proc. of the X Intern. conf. persp. techn. and meth. in MEMS design., 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 // *Матер. Міжнар. наук.-практ. конф. «Інформаційні управляючі системи та технології»*. – Одеса : Одеський національний 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 MEMSTECH2013: proc. of the IX Intern. conf. persp. techn. and meth. in MEMS design., 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 CSIT2012: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», Східноєвропейський національний ун-т ім. Лесі Українки. – Луцьк, 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. Гопяк // 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.
21. Tervoort B. Structurele analyse van visueel taalgebruik binnen een groep dove kinderen. [Structural Analysis of Visual Language Use within a Group of Deaf Children] / B. Tervoort. – Amsterdam : North Holland Press, 1953.

22. Stokoe W. C. Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf, Studies in linguistics: Occasional papers / W. C. Stokoe. – Buffalo : Dept. of Anthropology and Linguistics, University of Buffalo, 1960. – 78 p.
23. Farris M. Sign language research and Polish Sign Language / M. Farris // *Lingua Posnaniensis*. – 1994. – P. 13–36.
24. Боскис Г. М. Глухие и слабослышающие дети / Г. М. Боскис. – Москва : Изд-во АПН РСФСР, 1963. – 314 с.
25. Морозова Н. Г. Формирование познавательных интересов у аномальных детей / Н. Г. Морозова. – Москва : Просвещение, 1969. – 280 с.
26. Зайцева Г. Л. Жестовая речь. Дактилология / Г. Л. Зайцева. – Москва : Владос, 2000. – 192 с.
27. Краєвський Р. Г. Мова жестів глухих / Р. Г. Краєвський. – Київ : Рад. школа, 1964. – 219 с.
28. Адамюк Н. Б. Синтаксичні особливості УЖМ: на прикладі простого речення / Н. Б. Адамюк, І. І. Чепчина // *Жестова мова й сучасність : зб. наук. пр.* – Київ : Педагогічна думка, 2009. – Вип. 4. – С. 170–191.
29. Кульбіда С. В. Українська жестова мова як природна знакова система. Жестова мова й сучасність : зб. наук. пр. / С. В. Кульбіда. – Київ : Центродрук. – 2009. – С. 218–239.
30. Stone D. Interpreting in International Sign: Decisions of Deaf and non-deaf interpreters / D. Stone, D. Russel // Paper presented at the World Association of the Sign Language Interpreters in Durban. – South Africa, 2011.
31. Stein D. Sign Language Machine Translation Overkill / D. Stein, C. Schmidt, and H. Ney // In International Workshop on Spoken Language Translation (IWSLT). – Paris, 2010. – P. 337–344.
32. Проект «iCommunicator» [Електронний ресурс]. – Режим доступу: <http://www.icommunicator.com>.
33. Проект ViSiCAST [Електронний ресурс]. – Режим доступу: <http://www.visicast.co.uk/>.

34. Zhao L. A Machine Translation System from English to American Sign Language / L. Zhao, K. Kipper, W. Schuler, C. Vogler, N. Badler, M. Palmer // A In Envisioning Machine Translation in the Information Future: Proceedings of the Fourth Conference of the Association for Machine Translation (AMTA-00). – Cuernavaca, Mexico. –2000. – P. 293–300.
35. Проект Artemis [Электронный ресурс]. – Режим доступа: <http://www-sim.int-evry.fr/Artemis>.
36. Zijl V. L. A machine translation system for South African Sign Language/ D. Barker, V.L. Zijl // In Proceedings of Afrigraph 2003, ACM Siggraph Conference Proceedings Series. – Cape Town, South Africa. – 2003. – P. 49–52.
37. Suszczańska N. Continuous Text Translation using Text Modeling in the Thetos System / N. Suszczańska, P. Szmaj, S. Kulików // International Journal of Computational Intelligence. – № 4. – 2005. – P. 255–258.
38. Проект Sign Smith Studio [Электронный ресурс]. – Режим доступа: <http://www.vcom3d.com/?id=ssstudio>.
39. Проект SiSi [Электронный ресурс]. – Режим доступа: <http://mqtt.org/projects/sisi>.
40. Dreuw P. The SignSpeak Project – Bridging the Gap Between Signers and Speakers / P. Dreuw, H. Ney, G. Martinez, O. Crasborn, J. Piater, J. Miguel Moya, M. Wheatley // In Language Resources and Evaluation (LREC), Valletta, Malta, 2010. – P. 476–481.
41. Воскресенский А. Л. Компьютерный банк жестовой речи. // Компьютерная лингвистика и интеллектуальные технологии: Тр. Междунар. конф. Диалог'2003 (Протвино, 11–16 июня 2003 г.) / под ред. И. М. Кобозевой, Н. И. Лауфер, В. П. Селегея. – Москва : Наука, 2003. – С. 688–691.
42. Shoaib U. A platform-independent user-friendly dictionary from Italian to LIS / U. Shoaib, N. Ahmad, P. Prinetto, G. Tiotto // In: LREC'12 : 8th International Conference on Language Resources and Evaluation, Istanbul (Turkey). – 2012. – P. 2435–2438.

43. Morrissey S. An Example-based Approach to Translating Sign Language / S. Morrissey, A. Way // In Proceedings of the Workshop in Example-Based Machine Translation (MT Summit X), Phuket, Thailand. – 2005. – P. 109–116.
44. San-Segundo R. Evaluation of Alternatives on Speech to Sign Language Translation / R. San-Segundo, A. Pérez, D. Ortiz, L. F. D'Haro, M. I. Torres, F. Casacuberta // Proc. of Interspeech 2007. – Antwerp, Belgium. – 2007. – P. 2529–2532.
45. Dasgupta T. Prototype Machine Translation System From Text-To-Indian Sign Language / T. Dasgupta, S. Dandpat, A. Basu // In Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages. – 2008. – P. 19–26.
46. Morrissey S. Joining hands: developing a sign language machine translation system with and for the deaf community / S. Morrissey, A. Way // In Proceedings of the Conference and Workshop on Assistive Technologies for People with Vision and Hearing Impairments – Assistive Technology for All Ages (CVHI-07). – Granada, Spain.
47. Kanis J. Automatic Czech – Sign Speech Translation / J. Kanis, L. Müller // TSD 2007: Pilsen, Czech Republic. – 2007. – P. 488–495.
48. Stein D. Hand in Hand: Automatic Sign Language to Speech Translation / D. Stein, P. Dreuw, H. Ney, S. Morrissey, A. Way // In Proceedings of Theoretical and Methodological Issues in Machine Translation (TMI-07). – Sweden. – 2007. – P. 214–220.
49. Bungeroth J. Statistical Sign Language Translation / J. Bungeroth, H. Ney // Proc. LREC 2004, Workshop proceedings: Representation and Processing of Sign Languages. – Lisbon, Portugal. – 2004. – P. 105–108.
50. Almasoud A. M. A Proposed Semantic Machine Translation System for translating Arabic text to Arabic sign language / A. M Almasoud., H. S. Al-Khalifa // Journal of Software Engineering and Applications. – 2012. – Vol. 5, № 8. – P. 604–612.
51. Крак Ю. В. Реализация информационной технологии автоматизированного перевода с флективных языков на жестовый / Ю. В. Крак, А. В. Бармак, С. А. Романишин // Матер. IV Междунар. науч.-техн. конф. «Открытые

- семантические технологии проектирования интеллектуальных систем». – Минск, 2014. – С. 21–30.
52. Stokoe W. C. A dictionary of American sign languages on linguistic principles / W. C. Stokoe, C. C. Dorothy, G. C. Carl // Linstok Press, 1976. – 346 p.
53. Martin J. A linguistic comparison. The notation systems for signed languages: Stokoe Notation and Sutton SignWriting [Электронный ресурс]. – 2000. – Режим доступа: <http://www.signwriting.org/forums/linguistics/ling016.html>.
54. Система жестової нотації SignWriting [Электронный ресурс]. – Режим доступа: <http://www.signwriting.org/about/>.
55. Димскис Л. С. Изучаем жестовый язык : учеб. пособие для студентов дефектологических факультетов высших педагогических учебных заведений. – Москва : Издательский центр «Академия», 2002. – 128 с.
56. Prillwitz S. HamNoSys. Version 2.0: Hamburg Notation System for Sign Languages. An introductory guide / S. Prillwitz, T. Hanke. – Hamburg : Signum, 1989. – 46 p.
57. Словник Arbeitsgruppe Fachgebardenlexika, Fachgebardenlexikon Psychologic [Электронный ресурс]. – Hamburg, 1996. – Режим доступа: <http://www.sign-lang.uni-hamburg.de/PLEX/>.
58. Словник Johnston 1997 - 7'. Johnston. Royal N.S.W. Institute for deaf and blind children: Signs of Australia on CD-ROM: a dictionary of Auslan (Australian sign language). North Rocks. – 1997.
59. Kennedy G. A dictionary of New Zealand sign language. Auckland. – 1998.
60. Калустова О. Електронні бази паралельних та співвідносних текстів у перекладознавчих дослідженнях / О. Калустова // Вісн. Сум. держ. ун-ту. Сер. «Філол. науки». – № 1. – Суми, 2004. – С. 130–136.
61. Agris U. Towards a Video Corpus for Signer-Independent Continuous Sign Language Recognition / U. von Agris, K.-F. Kraiss // In Gesture in Human-Computer Interaction and Simulation. – Lisbon, Portugal. – 2007.

62. Vogler C. Handshapes and movements: Multiple-channel American Sign Language recognition / C. Vogler and D. N. Metaxas // In *Gesture Workshop*. – 2003. – P. 247–258.
63. Bowden. A Linguistic Feature Vector for the Visual Interpretation of Sign Language / R. Bowden, D. Windridge, T. Kadir, A. Zisserman, M. Brady // In *ECCV*. – Vol. 1. – 2004. – P. 390–401.
64. Bungeroth J. The ATIS Sign Language Corpus / J. Bungeroth, D. Stein, P. Dreuw, H. Ney, S. Morrissey, A. Way, L. van Zijl // In *Language Resources and Evaluation (LREC)*. – Marrakech, Morocco, 2008. – P. 2943–2946.
65. Stein D. Analysis of the German Sign Language Weather Forecast Corpus / D. Stein, J. Forster, U. Zelle, P. Dreuw, H. Ney // In *4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010)*. – Valletta, Malta, 2010. – P. 225–230.
66. Bungeroth J. A german sign language corpus of the domain weather report / J. Bungeroth, D. Stein, P. Dreuw, M. Zahedi, H. Ney // In *Fifth International Conference on Language Resources and Evaluation*. – Italy, 2006. – P. 2000–2003.
67. Lombardo V. An Avatar-based Interface for the Italian Sign Language / V. Lombardo, C. Battaglini, R. Damiano, F. Nunnari // *International Conference on Complex, Intelligent and Software Intensive Systems*. – Seoul, Korea, 2011. – P. 590–594.
68. Бармак О. Комп'ютерна технологія для перекладу з української на жестову мову / О. Бармак, С. Романишин // *Вісник ТНТУ*. – 2012. – Т. 65. – № 1. – С. 195–208.
69. Кобрицов Б. П. Снятие семантической многозначности глаголов с использованием моделей управления, извлеченных из электронных толковых словарей / Б. П. Кобрицов, О. Н. Ляшевская, С. Ю. Толдова [Электронный ресурс]. – Режим доступа: <http://download.yandex.ru/IMAT2007/kobricov.pdf>.
70. Романюк А. Б. Методи вирішення лексичної багатозначності. Використання WORDNET для вирішення проблем багатозначності / А. Б. Романюк, І. А. Сундутова, М. М. Романишин // *Вісник Нац. ун-ту «Львів. політехніка»*

«Комп'ютерні системи проектування. Теорія і практика». – 2011. – № 711. – С. 147–157.

71. Вороной А. С. Використання онтологій для підвищення якості пошуку інформації для поповнення баз знань інтелектуальних систем // Матер. Міжнар. наук.-техн. конф. «Комп'ютерні науки і інженерія 2009», (Львів, 14–16 травня 2009 р.). – Львів, 2009. – С. 364–366.
72. Brochhausen M. Establishing and Harmonising Ontologies in an Interdisciplinary Health Care and Clinical Research Environment. – 2007. – [Електронний ресурс]. – Режим доступу: <http://ontology.buffalo.edu/medo/eHealth.pdf>.
73. Kumar A. The Unified Medical Language System and the Gene Ontology: Some Critical Reflections / A. Kumar, B. Smith // In: KI 2003: Advances in Artificial Intelligence (Lecture Notes in Artificial Intelligence 2821). – Berlin : Springer, 2003. – P. 135–148.
74. Проект UNSPSC [Електронний ресурс]. – Режим доступу: unspsc.org.
75. Al-Halimi R. WordNet: An Electronic Lexical Database / R. Al-Halimi, R. C. Berwick, J. F. M. // Burg Edited by Christiane Fellbaum. – Cambridge, MA : MIT Press, 1998. – 422 p.
76. Проект ConceptNet [Електронний ресурс]. – Режим доступу: <http://conceptnet5.media.mit.edu/>.
77. Проект PlWordNet [Електронний ресурс]. – Режим доступу: <http://plwordnet.pwr.wroc.pl/wordnet/>.
78. Chu-Ren H.. Chinese WordNet: Design and Implementation of a Cross-Lingual Knowledge Processing Infrastructure / H. Chu-Ren, H. Shu-Kai, H. Jia-Fei, C. Yun-Zhu, S. I-Li, C. Yong-Xiang, H. Sheng-Wei // Journal of Chinese Information Processing. – 2010. – № 24(2). – P. 14–23.
79. Benoît S. Building a free French wordnet from multilingual resources / S. Benoît, F. Darja // In Proc. of Ontolex 2008. – Marrakech, Maroc, 2008.
80. Pianta E. MultiWordNet: Developing an aligned multilingual database / E. Pianta, L. Bentivogli, C. Girardi // In Proc. of the 1st International Conference on Global WordNet. – Mysore, India, 2002. – P. 21–25.

81. Tufis D. Balkanet: Aims, methods, results and perspectives. A general overview / D. Tufis, D. Cristea, S. Stamou // Romanian J. Sci. Tech. Inform. (Special Issue on Balkanet). – 2004. – № 7(1–2). – P. 9–43.
82. Henrich V. Semi-Automatic Extension of GermaNet with Sense Definitions from Wiktionary / V. Henrich, E. Hinrichs, T. Vodolazova // In: Proceedings of 5th Language & Technology Conference. – Poznań, Poland, 2011. – P. 126–130.
83. Bhattacharyya P. IndoWordNet / P. Bhattacharyya // Lexical Resources Engineering Conference 2010 (Malta, May). – 2010.
84. Проект RussNet [Електронний ресурс]. – Режим доступу: http://www.phil.pu.ru/depts/12/RN/index_ru.shtml.
85. Кульчицький І. М. Розроблення wordnet-подібного словника української мови / І. М. Кульчицький, А. Б. Романюк, Х. Б. Харів // Вісник Нац. ун-ту «Львівська політехніка»: Інформаційні системи та мережі. – 2011. – № 673. – С. 306–319.
86. Анісімов А. В. UWN: Універсальна онтологічна база знань української мови / А. В. Анісімов, О. О. Марченко, А. О. Никоненко // Проблеми програмування. – 2012. – № 2–3. – С. 348–355.
87. Balkova V. Russian WordNet. From UML-notation to Internet/Intranet Database Implementation / V. Balkova, A. Suhonogov, S.A. Yablonsky // Proceedings of the Second International WordNet Conference, Brno, Czech Republic. – 2004. – P. 31–38.
88. Quadros R. M. Clause structure. Sign Languages: A Cambridge Language Survey / R. M. Quadros, D. Lillo-Martin // In Diane Brentari (Ed.), Cambridge University Press, 2010. – P. 225–251.
89. Milković M. Word order in Croatian Sign Language / M. Milković, S. Bradarić-Jončić, R. Wilbur // Investigating Understudied Sign Languages – Croatian SL and Austrian SL, with comparison to American SL, special issue of Sign Language & Linguistics 9:1/2, 2006. – P. 169–206.
90. Lin X. Study on Translating Chinese into Chinese Sign Language / X. Lin, J. Wen // Computer Science & Technology. – 2010. – Vol. 15, № 5. – P. 485–490.

91. Воскресенский А. Л. О распознавании жестов языка глухих / А. Л. Воскресенский, С. Н. Ильин, М. Zelezny // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Междунар. конф. «Диалог» (Бекасово, 26–30 мая 2010 г.). – Москва : РГГУ, 2010. – Вып. 9 (16). – С. 76–81.
92. Гриф М. Г. Проблема автоматизации сурдоперевода с позиции прикладной лингвистики / М. Г. Гриф, М. К. Тимофеева // Сибирский филологический журнал. – 2012. – № 1. – С. 211–219.
93. Лурия А. Р. Понимание компонентов речевого высказывания // Язык и сознание / Ред. Е. Д. Хомской. – Москва : Изд. МГУ, 1979. – С. 217–226.
94. San-Segundo R. Speech to sign language translation system for spanish / R. San-Segundo, R. Barra, R. Córdoba, L. F. D'Haro, F. Fernández, J. Ferreiros, J. M. Lucas, J. Macías-Guarasa, J. M. Montero and J.M. Pardo // Speech Communication. – 2008. – Vol. 50. – P. 1009–1020.
95. Проект «Spreadthesign» [Электронный ресурс]. – Режим доступа: <http://www.spreadthesign.com/ua/>.
96. Берко А. Ю. Системи баз даних та знань. Книга 1. Організація баз даних та знань : навч. посіб. / А. Ю. Берко, О. М. Верес, В. В. Пасічник. – Львів : Магнолія 2006, 2008. – 456 с.
97. Вихованець І. Р. Граматика української мови. Синтаксис / І. Р. Вихованець. – Київ, 1993. – 368 с.
98. Проект «UkrParser» [Электронный ресурс]. – Режим доступа: <https://github.com/mdavydov/UkrParser>.
99. Navigli R. Word Sense Disambiguation: A Survey / R. Navigli // ACM Computing Surveys, ACM Press, 2009. – № 41(2). – P. 1–69.
100. Захарова И. В. Способы автоматического построения онтологии для задач анализа текстов / И. В. Захарова, М. С. Тимченко // Знання – Онтології – Теорії : тр. Всеросійск. конф. ЗОНТ-09. – Новосибірск, 2009. – Т. 2. – С. 164–167.

101. Антонов И. В. Модель онтологии предметной области для систем семантически-ориентированного доступа / И. В. Антонов // Тр. Псковского политехнического института. Электротехника. Машиностроение. – 2011. – № 14.3. – С. 339–343.
102. Проект DOE [Электронный ресурс]. – Режим доступа: <http://opales.ina.fr/public/index.html> Farquhar.
103. Fickas A. The Ontolingua Server: a tool for collaborative ontology construction / A. Fickas, J. Rice // Proceedings of the 10th Banff Knowledge Acquisition for Knowledge based system workshop. – 1995.
104. Sure Y. OntoEdit: Collaborative Engineering of Ontologies / Y. Sure, M. Erdmann, R. Studer // Proceedings of the First International Semantic Web Conference 2002, Sardinia, Italia, 2002. – P. 221–235.
105. Domingue J., Motta E., Garcia C. O. Knowledge modelling in WebOnto and OCML: A user guide / J. Domingue, E. Motta, C. O. Garcia. – 1999.
106. Noy N. F. Creating Semantic Web Contents with Protege-2000 / N. F. Noy, M. Sintek, S. Decker, M. Crubez, R. W. Ferguson, M. A. Musen // IEEE Intelligent Systems, 2011. – № 16. – P. 60–71.
107. Казакова А. С. Методы и инструменты реализации предметно-ориентированных языков программирования / А. С. Казакова. – Москва : Вильямс, 2009. – С. 1–79.
108. Фаулер М. Предметно-ориентированные языки программирования : пер. с англ. / М. Фаулер. – Москва : ООО «И.Д. Вильямс», 2011. – 576 с.
109. JetBrains MPS, MetaProgramming System [Электронный ресурс]. – Режим доступа: <http://www.jetbrains.com/mps/>.
110. IntelliJ IDEA, code-centric IDE [Электронный ресурс]. – Режим доступа: <http://www.jetbrains.com/idea/>.
111. Xtext, framework for development of programming languages and domain specific languages [Электронный ресурс]. – Режим доступа: <http://www.eclipse.org/Xtext/>.

112. Жуков М. В. Реализация предметно-ориентированных языков средствами фреймворка Xtext // Науч. журнал Известия Пензенского государственного педагогического унив. им. В. Г. Белинского, Физико-математические и технические науки. – 2009. – № 13 (17). – С. 98–102.
113. Білодід І. К. Словник української мови. – Київ : Наукова думка, 1970–1980. – Т. 1–11.
114. Давидов М. В. Програмний тренажер навчання сурдоперекладу / М. В. Давидов, В. В. Пасічник, Ю. В. Нікольський, С. М. Тиханський // Проблеми освіти : наук.-метод. зб. – 2009. – № 60. – С. 108–113.
115. Загнітко А. П. Основи українського теоретичного синтаксису. Ч. 1 / А. П. Загнітко. – Горлівка, 2004. – 227 с.
116. Дарчук Н. П. Автоматичний синтаксичний аналіз текстів корпусу української мови / Н. П. Дарчук // Українське мовознавство. – Київ. нац. ун-т ім. Т. Шевченка, 2013. – № 43. – С. 11–19.
117. Koehn P. Machine Translation Systems for Europe / P. Koehn, A. Birch, R. Steinberger // Machine Translation Summit XII. – 2009. – P. 65–72.
118. Othman A. Statistical Sign Language Machine Translation: from English written text to American Sign Language Glos / A. Othman, M. Jemni // IJCSI International Journal of Computer Science Issues. – Vol. 8, Issue 5, № 3. – 2011. – P. 65–73.
119. Goyal P. Converting Phrase Structures to Dependency Structures in Sanskrit / P. Goyal, A. Kulkarni // Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, Dublin, Ireland, 2014. – P. 1834–1843.

ДОДАТОК А

**Корпус паралельних текстів «Українська словесна мова –
Українська жестова мова»**

- | | |
|---|------------------------------------|
| 1. Як тебе звати? | Ти ім'я? |
| 2. Ти є нечуючий? | Нечуючий ти? |
| 3. Ти є студент? | Студент ти? |
| 4. Чи розумієш ти його / її? | Ти розуміти його / її? |
| 5. Хто він / вона є? | Хто він /вона? |
| 6. Повтори як тебе звати? | Повтори, ім'я ти? |
| 7. Вона чи він є студентом? | Студент вона /він? |
| 8. Це є твоє? | Це твій? |
| 9. Ти є чуючою людиною? | Чуючий ти? |
| 10. Ви є вчителем? | Вчитель ви? |
| 11. Де ти вивчаєш жести? | Ти вчити жести, де? |
| 12. Чому ти вивчаєш жести? | Ти вчити жести, чому? |
| 13. Як ви покажете жест «дякую»? | Як жест д-я-к-у-ю? |
| 14. Вони вивчають жестову мову? | Вони вчити жести? |
| 15. Хто твій вчитель? | Твій вчитель, хто? |
| 16. Чи тобі подобається вивчати жестову мову? | Ти подобається вчити жестова мова? |
| 17. Чи розумієте ви його / її? | Ти розуміти він / вона, ти? |
| 18. Чи подобається вам зустрічатися з нечуючими людьми? | Ти подобається зустріч нечуючий? |
| 19. У вас є діти? | Діти ти? |
| 20. Скільки в тебе є сестер? | Сестра скільки ти? |
| 21. Як звати твою маму? | Твоя мама ім'я як? |
| 22. Твій батько глухий? | Твій батько глухий? |
| 23. Де ти працюєш? | Ти робота де? |
| 24. Де ти живеш? | Ти жити де? |
| 25. Тебе звати Петро? | Ти ім'я П-е-т-р-о, ти? |
| 26. Ви розлучилися? | Ви розлучитися, ви? |
| 27. Твоя сестра одружена? | Твоя сестра одружена? |
| 28. Ти зустрічався з моїм братом? | Ти зустрічати мій брат, ти? |
| 29. В якому місті ти живеш? | Місто ти жити? |
| 30. Тобі подобається вивчати жести? | Ти подобається вчити жести? |
| 31. Ваш будинок великий? | Ваш будинок великий? |
| 32. Скільки у вас дітей? | Діти скільки ви? |
| 33. Скільки кімнат у вашому будинку? | Ваш будинок, скільки кімнат? |
| 34. Тобі подобається твоя робота? | Ти подобається твоя робота? |
| 35. Ти тут живеш? | Ти жити тут? |
| 36. Ти ходиш до школи? | Ти ходити школа ти? |
| 37. В якому місті ви працюєте? | Місто ви працювати? |

- | | |
|---|--|
| 38. Ви одружені? | Одружені ви? |
| 39. Тобі подобаються малі міста? | Мале місто, ви подобатися? |
| 40. Ви хочете зустрітися з моєю мамою? | Хотіти ви зустріти моя мама ви? |
| 41. Вона тобі подобається? | Ти подобатися вона? |
| 42. Ти самотній? | Ти самотній ти? |
| 43. Мої тітка і дядько розлучені. | Мій тітка дядько розлучені. |
| 44. Тітка любить дядька. | Тітка любити дядька. |
| 45. Тітка не хотіла розлучатись. | Тітка не хотіти розлучатися. |
| 46. Вона є хорошим другом. | Вона друг хороший. |
| 47. Який твій улюблений колір? | Твій улюблений колір який? |
| 48. Вам подобається коричневий колір? | Ви подобається колір
коричневий? |
| 49. Скільки тобі років? | Років ти? |
| 50. Ти хочеш йти додому зараз? | Ти хотіти йти додому зараз? |
| 51. Якого кольору ваш папір? | Ваш папір який колір? |
| 52. Тобі подобаються червоні автомобілі? | Ти подобається червоний
автомобіль? |
| 53. Коли ви переїхали сюди? | Ви їхати сюди коли? |
| 54. Який вид цукерок вам подобається? | Цукерка ви подобається який
вид? |
| 55. Яка твоя улюблена страва? | Твоя улюблена їжа яка? |
| 56. Який суп тобі подобається? | Суп ти подобається який? |
| 57. Ти голодний? | Голодний ти? |
| 58. Ти любиш їсти апельсини? | Апельсин ти подобається їсти? |
| 59. Чи є у тебе рюкзак? | Рюкзак ти мати? |
| 60. Коли ви змінюєте одяг? | Ви міняти одяг коли? |
| 61. Якого кольору халат лікаря? | Лікар халат колір який? |
| 62. Хто в цьому номері має окуляри? | Хто ця кімната мати окуляри? |
| 63. Якого кольору твоя сорочка? | Твоя сорочка який колір? |
| 64. Чи є у вашій ванній кімнаті ванна? | Ваша ванна кімната мати ванну? |
| 65. Чи є у тебе автомобіль? | Машину мати ти? |
| 66. Якою зубною пастою ти користуєшся? | Зубна паста ти користуватись
який? |
| 67. У вашій спальні є вікно? | Ваша спальня мати вікно? |
| 68. Чи люблять коти їсти пташок? | Кіт любити їсти пташка? |
| 69. Які книги ви любите читати? | Книга ви любити читати, який? |
| 70. Де ми можемо купити молоко? | Молоко купити де? |
| 71. Ти хочеш мати коня? | Кінь ти хотіти? |
| 72. У мене є папуга. | Я мати папуга. |
| 73. Розкажіть мені, як ви себе почуваете. | Розкажи мені як ви почуватися. |
| 74. У мене диван коричневого кольору. | Я мати диван колір коричневий. |
| 75. Наш холодильник білого кольору. | Наш холодильник колір білий. |
| 76. Моя мама пере кожного дня. | Моя мама прати день кожен. |
| 77. Іван має 20 пар взуття. | І-в-а-н мати 20 взуття. |

78. Я добре маюю. Я малювати добре.
79. Я люблю український борщ. Я любити борщ український.
80. Я дуже люблю піцу. Я любити піца дуже.
81. Він не любить цукерок. Він не любити цукерка.
82. Я подивилася цікавий фільм. Я дивитися фільм цікавий закінчила.
83. Я піду завтра до школи. Я йти школа завтра.
84. Я живу тут 10 років. Я жити тут 10 років.
85. Радий тебе бачити. Радий зустрічатися ти.
86. Я можу розмовляти жестовою мовою небагато. Жестова мова мало я можу.
87. Пробачте, я не розумію це. Пробачте, я розуміти не ти.
88. Говоріть повільніше, будь ласка. Говорити повільно будь ласка.
89. Я маю коричневий ремінь. Я мати ремінь коричневий.
90. В полі росте велике дерево. Поле дерево великий рости.
91. Люди багато працюють. Люди працювати багато.
92. Я живу в місті. Я жити місто.
93. Я люблю їсти червоні яблука. Я любити їсти яблука червоний.
94. Мене звати Іван. Я звати Іван.
95. Я глухий. Я глухий.
96. Я знаю жестову мову добре. Я жести знати добре.
97. Я зрозумів тебе добре. Я ти розуміти добре.
98. Мій батько глухий а мати слабочуюча. Батько глухий чисто мати слабочуюча.
99. Мої дідусь і бабуся чуючі. Мій дідусь бабуся говорити чисто.
100. Моя бабуся має зелене пальто. Моя бабуся мати пальто зелений.
101. Мої батьки знають жестову мову. Мої батьки жести знати.
102. Моя сестра любить ходити до школи. Моя сестра любити школа.
103. Я купила хліб. Я купити вже хліб.
104. Смачний пиріг на тарілці. Тарілка пиріг смачний.
105. На тарілці мало пирогів. Тарілка пиріг мало.
106. Я не хочу тебе бачити. Я бачити тебе не хочу.
107. Я не працюю. Я працюю не.
108. Сьогодні банкету не буде. Банкет сьогодні скасовується.
109. У дівчини плаття немодне. Плаття дівчини немодне.
110. Маленька дівчинка годує жовтих курчат. Дівчинка маленький годувати курчата жовтий.
111. Мама миє вікно ганчіркою. Мама вікно мити ганчірка.
112. Тканина загорілась від перегріву праски. Праска довго тканина горіти.
113. На зеленому дереві лежить сніг. Дерево зелений сніг лежати.
114. Я вивчила уроки. Я вчити уроки вже.
115. Я бігала вчора. Я бігати вчора.
116. Я не люблю грати у теніс. Теніс я любити грати не.

117. У мене немає собаки. Собака я мати не.
 118. Я маю коричневого собаку. Я мати собака коричневий.
 119. Я зайшла додому тихо. Дім я тихо входити.
 120. Я йду в магазин о 9.00 ранку. 9 годин зранку магазин я йти.
 121. Я можу піти в магазин замість тебе. Для ти магазин я йти можна.
 122. Магазин закривається о 8 вечора. Магазин закрити 8 вечір.
 123. Я люблю плавати, але не люблю бігати. Плавати я любити, бігати я любити не.
 124. У мене є три коти. Я мати кіт три.
 125. Я їду о 5 годині. Я їхати година 5.
 126. Я йду до ресторану. Я йти ресторан зараз.
 127. Я бачила твою маму вчора. Вчора я бачити мама ти.
 128. Я пообідав. Я обід вже.
 129. Я не вірю тобі. Я ти вірити не.
 130. Вона є моїм начальником. Вона мій начальник.
 131. Я маю два брата і три сестри. Я мати брат 2 сестри 3.
 132. Він дав їй книжку. Він давати вона книжка.
 133. Володя купив машину. В-о-л-о-д-я купити машина.
 134. Чашка впала на підлогу. Чашка падати підлога.
 135. У неї висока температура. Вона мати температура високий.
 136. Він любить кататися на лижах. Він любити лижі.
 137. Сьогодні гарна погода. Погода гарний сьогодні.
 138. Дорога йде рівно. Дорога рівно.
 139. Хлопець йде додому. Хлопець йти дім.
 140. Час йде швидко. Час летить.
 141. Я довіряю людям. Я довіряти людина людина.
 142. Я довіряла людям. Я довіряти був людина людина.
 143. Я довірятиму людям. Я довіряти буду людина людина.
 144. Я маю собаку давно. Я мати давно собака.
 145. Я виходжу з дому. Я виходити дім.
 146. Ти маєш собаку? Ти мати собака?
 147. Я не маю собаку давно. Я мати не собака давно.
 148. Сьогодні негарна погода. Сьогодні погода поганий.
 149. Я не виходжу з дому. Я виходити не дім.
 150. Я дивлюсь цікавий фільм. Я дивитися фільм цікавий.
 151. Я подивилася цікавий фільм. Я дивитися вже фільм цікавий.
 152. Серце б'ється швидко. Серцебиття.
 153. Я читаю цікаву книжку. Я читати книга цікавий.
 154. Я прочитала цікаву книжку. Я читати вже книга цікавий.
 155. Я несу рюкзак. Я нести_рюкзак.
 156. Я несу сумку. Я нести_сумка.
 157. Я не маю меблів у своєму будинку. Я не_маю стіл стілець ліжко будинок мій.
 158. Я люблю пташок. Я любити пташка пташка.

159. Він має цукерки.
 160. Я бачу зелені дерева.
 161. Він увімкнув комп'ютер.
 162. Мама заплітає мені коси.
 163. Марічка помила посуд.
 164. Сонечко посміхається мені.
 165. Я вперше побачила слона.
 166. У зоопарку - санітарний день.
 167. Я танцювала на сцені.
 168. У мене сьогодні день народження.
169. Концерт закінчився.
 170. Ми зайшли у цукерню.
 171. Бабуся насмажила млинців.
 172. Бабуся смажить картоплю.
 173. У мене два місяці відпустки.
 174. Вона випила кави.
 175. Я заплакала.
 176. Мені не хочеться їсти.
 177. Ми разом снідаємо.
 178. Я сама купую ягоди.
 179. Ти вмієш співати?
 180. Я була гарною?
 181. Я була зайнята.
 182. Він був голодний сьогодні.
 183. Тобі було цікаво?
 184. Хто сьогодні черговий?
 185. Де росте кавун?
 186. Хто живе в дуплі?
 187. Чия це нора?
 188. Який у тебе розмір ноги?
 189. Що ти любиш їсти?
 190. Де тато?
 191. Ти боїшся холоду?
 192. Літак готовий до посадки.
 193. Я прокидаюся рано.
 194. Ця хвороба заразна.
 195. Цей хлопець сповнений енергії.
 196. Моя дочка ходить у дитячий садок.
197. Ми їдемо у автобусний тур.
 198. Я стою на автобусній зупинці.
 199. Я буду стояти на своєму.
 200. Вчора йшов дощ.

- Він має цукерки.
 Я бачити багато дерево зелений.
 Він включити комп'ютер.
 Мама плести_коса мені.
 Марічка мити вже посуд.
 Сонце посміхатися мені.
 Я бачити слон перший раз.
 Зоопарк – санітарний день
 Я танцювати вже сцена.
 Я мати день_народження
 сьогодні.
- Концерт завершився.
 Ми зайти цукерня.
 Бабуся смажити вже млинці.
 Бабуся смажити_картопля.
 Я мати два місяці відпустка.
 Вона пити готово кава.
 Я плакати вже.
 Я їсти не_хотіти.
 Ми снідати разом.
 Я купляти ягоди сам.
 Ти співати?
 Гарна була я?
 Я бути_зайнятим.
 Він бути_голодним сьогодні.
 Було цікаво ти?
 Черговий сьогодні хто?
 Кавун рости де?
 Жити дупло хто?
 Ця нора чия?
 Твій розмір нога який?
 Ти подобається їсти що?
 Тато де?
 Холод боятися ти?
 Літак готовий посадка.
 Я прокинутися рано.
 Ця хвороба заразний.
 Цей хлопець енергійний.
 Моя дочка ходити
 дитячий_садок.
 Ми їхати автобусний_тур.
 Я стояти автобусна_зупинка.
 Я бути стояти_на_своєму.
 Вчора йти_дощ.

ДОДАТОК Б

Файли граматично доповненої онтології

category ім_наз

category ім_род

category ім_дав

category ім_знах

category ім_оруд

category ім_місц = іменник у місцевому відмінку (на кому на чому?)

category д_інф = дієслово (інфінітив)

category д_жест = дієслово (жест)

category ім_жест = іменник (жест)

synset носій_знань

synset містить_знання

synset читати_засвоювати

synset читати_декламувати

synset слухати_засвоювати

synset набувати_знання

synset числити_визначати

synset сприймати_розуміти

synset кількість_предметів

synset поставлена_проблема

synset музичний_інструмент

synset інструмент

synset площина

synset графічний_знак

synset сутність

synset людина_особа

synset предмет

synset вид_заняття

synset те_що_можна_співати

synset музичний_твір

synset учити_навчати = передавати знання

учити 1.ім_знах 2.ім_дав

давати поради 1.ім_дав з 2.ім_род

1: носій_знань

2: містить_знання

synset учити_вивчати = набувати певних знань

учити 1.ім_знах

засвоювати 1.ім_знах

опановувати 1.ім_знах

запам'ятовувати 1.ім_знах

вивчати 1.ім_знах

1: містить_знання

synset вчитися_у = набувати певних знань у певному закладі

вчитися в 1.ім_місц

відвідувати 1.ім_знах

жест вчитися 1.ім_жест

1: навчальний_заклад

synset рахувати_лічити = визначати кількість

рахувати 1.ім_знах

рахувати на 1.ім_місц
лічити 1.ім_знах
лічити на 1.ім_місц
 1: кількість_предметів

synset розв'язувати_вирішувати = знаходити відповідь
 розв'язувати 1.ім_знах
вирішувати 1.ім_знах
 1: поставлена_проблема

synset писати_записувати = зображувати рукою графічні знаки
писати 1.ім_знах
писати на 2.ім_місц
писати у 2.ім_місц
писати про 3.ім_знах
записувати 1.ім_знах
записувати на 2.ім_місц
записувати у 2.ім_місц
записувати про 3.ім_знах
 1: графічний_знак
 2: площина
 3: сутність

synset говорити_розповідати = висловлювати думки
говорити 1.ім_дав 2.ім_знах
говорити 1.ім_дав про 2.ім_знах
говорити до 1.ім_знах
 розповідати 1.ім_дав 2.ім_знах
 розповідати 1.ім_дав про 2.ім_знах
 1: носій_знань
 2: містить_знання

synset читати_прочитувати = сприймати що-небудь записане літерами
 читати 1.ім_дав 2.ім_знах
 читати 1.ім_дав про 2.ім_знах
 читати 1.ім_дав з 3.ім_знах
 прочитувати 1.ім_дав 2.ім_знах
 прочитувати 1.ім_дав про 3.ім_знах
 прочитувати 1.ім_дав з 2.ім_знах
 1: носій_знань
 2: містить_знання
 3: сутність

synset займатися_оволодівати = здійснювати що-небудь
 займатися 1.ім_оруд
 оволодівати 1.ім_оруд
 1: вид_заняття

synset співати_виспівувати = виконувати голосом музичний твір
 співати 1.ім_дав 2.ім_знах
 співати 1.ім_дав про 3.ім_знах
 виспівувати 1.ім_дав 2.ім_знах
 виспівувати 1.ім_дав про 3.ім_знах
 1: носій_знань
 2: те_що_можна_співати
 3: сутність

synset грати_вигравати = відтворювати що-небудь на музичному інструменті
 грати 1.ім_дав 2.ім_знах на 3.ім_місц
 вигравати 1.ім_дав 2.ім_знах на 3.ім_місц
 1: носій_знань

2: музичний_твір
 3: музичний_інструмент

synset малювати_зображати = зображувати на площині
 малювати 1.ім_знах на 2.ім_місц 3.ім_оруд
 зображати 1.ім_знах на 2.ім_місц 3.ім_оруд
 1: предмет
 2: площина
 3: інструмент

synset підручник_книга
 підручник
 synset том_книга
 том
 synset буквар_книга
 synset абетка_книга
 synset журнал
 synset задачник_книга
 synset збірник_книга
 synset книга
 synset видання_книга
 synset зошит
 synset щоденник
 synset ручка
 synset олівець
 synset сторінка
 synset номер_сторінки
 synset студент
 synset університет
 synset школа
 synset клас
 synset навчання
 synset учень
 synset школяр
 synset урок
 synset навчальний_предмет навчальний предмет
 synset математика
 synset читання
 synset українська_мова українська мова
 synset англійська_мова англійська мова
 synset музика
 synset малювання
 synset фізкультура
 synset задача
 synset число
 synset літера
 synset слово
 synset словосполучення
 synset речення
 synset текст
 synset вірш
 synset спорт
 synset пісня
 synset спів
 synset група_в_університеті група
 synset навчальний_заклад навчальний заклад

common(університет) = навчальний_заклад
 common(школа) = навчальний_заклад
 common(школяр) = учень

```

common(математика) = навчальний_предмет
common(малювання) = навчальний_предмет
common(англійська_мова) = навчальний_предмет
common(українська_мова) = навчальний_предмет
common(фізкультура) = навчальний_предмет
common(читання) = навчальний_предмет
common(музика) = навчальний_предмет
common(книга) = носій_знань
common(підручник_книга) = книга
common(задачник_книга) = книга
common(том_книга) = книга
common(підручник_книга) = книга
common(книга) = носій_знань
common(учити_вивчати) = набувати_знання
common(пісня) = те_що_можна_співати
common(те_що_можна_співати) = музичний_твір
common(людина_особа) = сутність
common(предмет) = сутність
common(літера) = графічний_знак
common(олівець) = інструмент
common(ручка) = інструмент

parts(слово) = літера
parts(словосполучення) = слово
parts(речення) = слово
parts(текст) = речення
parts(вірш) = слово
parts(пісня) = слово
parts(школа) = клас
parts(фізкультура) = спорт
parts(навчальний_предмет) = урок
parts(книга) = сторінка // сторінка є меронімом книги. книга є холонімом сторінки,

optional_parts(сторінка) = номер_сторінки
optional_parts(учити_вивчати) = читати_засвоювати, слухати_засвоювати,
сприймати_розуміти
optional_parts(рахувати_лічити) = числити_визначати
optional_parts(читати_прочитувати) = сприймати_розуміти
optional_parts(навчання) = урок
optional_parts(вид_заняття) = навчальний_предмет

associations(студент) = університет, група_в_університеті
associations(школяр) = школа, клас
associations(фізкультура) = спорт
associations(пісня) = музика
associations(текст) = читання
associations(вірш) = читати_декламувати
associations(пісня) = співати_виспівувати
associations(задача) = розв'язувати_вирішувати
associations(школяр) = клас
associations(слово) = писати_записувати
associations(текст) = читати_прочитувати
associations(фізкультура) = займатися_оволодівати
associations(число) = рахувати_лічити
associations(навчальний_предмет) = учити_вивчати
associations(книга) = читати_прочитувати
associations(ручка) = писати_записувати

```

ДОДАТОК В

Акти про впровадження результатів дисертаційних досліджень

**ЛЬВІВСЬКА СПЕЦІАЛЬНА ЗАГАЛЬНООСВІТНЯ
ШКОЛА-ІНТЕРНАТ МАРІЇ ПОКРОВИ ДЛЯ ГЛУХИХ ДІТЕЙ**

А К Т
впровадження результатів дисертаційної роботи
“Система комп’ютерного перекладу української жестової мови з
використанням граматично доповненої онтології”
Лозинської Ольги Володимирівни,
асистента кафедри інформаційних систем та мереж
Національного університету “Львівська політехніка”

Цей акт підтверджує, що основні теоретичні та практичні результати дисертаційної роботи Лозинської О.В. за темою “Система комп’ютерного перекладу української жестової мови з використанням граматично доповненої онтології”, представленої на здобуття наукового ступеня кандидата технічних наук, використовуються у Львівській спеціальній загальноосвітній школі-інтернаті Марії Покрови для глухих дітей шляхом впровадження програмно-алгоритмічного комплексу «Перекладач української жестової мови».

Програмно-алгоритмічний комплекс «Перекладач української жестової мови» встановлено на сервер школи, передано повний комплект програмної документації, інструкцію користувача та адміністратора системи.

Розроблена Лозинською О.В. система комп’ютерного перекладу дозволяє здійснювати переклад речень української словесної мови у речення анотованої української жестової мови, відтворювати відео перекладених жестів, а також дає змогу здійснювати синтаксичний аналіз речень української словесної мови та речень української жестової мови.

Створені Лозинською О.В. у процесі дослідження методи та програмні засоби опрацювання та перекладу української жестової мови використовуються для задоволення освітніх, розважальних та соціо-комунікативних потреб, які дають можливість отримання знань та вмінь, які необхідні у навчальному процесі Львівської спеціальної загальноосвітньої школи-інтернату Марії Покрови для глухих дітей.

Директор Львівської спеціальної
загальноосвітньої школи-інтернату
Марії Покрови для глухих дітей



Ю.М. Крамар

Завідувач кафедри інформаційних
систем та мереж, д.т.н.

В.В. Литвин

м. Львів

А К Т

**впровадження результатів дисертаційного дослідження
Лозинської Ольги Володимирівни за темою
“Система комп’ютерного перекладу української жестової мови з використанням
граматично доповненої онтології”**

Цей акт підтверджує, що результати дисертаційного дослідження Лозинської Ольги Володимирівни за темою “Система комп’ютерного перекладу української жестової мови з використанням граматично доповненої онтології” використовуються у ПП Фірма “Біокурс” для підготовки та створення матеріалів, анотованих українською жестовою мовою.

Програмно-алгоритмічний комплекс перекладу української жестової мови, співавтором якого є Лозинська О.В., дає змогу здійснювати автоматичний переклад української словесної мови на українську жестову мову, переглядати відео перекладених жестів, а також дає змогу здійснювати семантично-синтаксичний аналіз речень української словесної мови та речень української жестової мови.

У програмно-алгоритмічному комплексі використовуються такі результати дисертаційного дослідження Лозинської О.В.: система правил семантично-синтаксичного розбору речень для української словесної та анотованої української жестової мов, що дало змогу здійснити семантично-синтаксичний аналіз речень для обох мов, граматично доповнена онтологія української словесної та анотованої української жестової мов, яка використовується для побудови системи правил перекладу анотованої української жестової мови.

Програмно-алгоритмічний комплекс є зручним та доступним як для чуючих, так і для нечуючих користувачів з навичками роботи на комп’ютері, та дає змогу комфортно реалізувати двосторонній переклад УЖМ. Дисертантом складено та сформовано повний комплект відповідної програмної документації та інструкції по користуванню системою.

Впровадження програмно-алгоритмічного комплексу перекладу української жестової мови дало змогу реалізувати комфортний доступ до сучасних інформаційних ресурсів, поданих в інформаційних системах для осіб з вадами слуху.

Директор ПП Фірма “Біокурс”



Т.В. Коваль

“ 25 ” серпня 2015 р.



«Затверджено»
 Проректор з науково-педагогічної роботи
 Національного університету
 «Львівська політехніка»
 О.Р. Давидчак
 2016 р.

про впровадження в навчальний процес результатів
 кандидатської дисертаційної роботи
Лозинської Ольги Володимирівни

Цей акт складено про те, що результати кандидатської дисертаційної роботи Лозинської Ольги Володимирівни на тему «Система комп'ютерного перекладу української жестової мови з використанням граматично-доповненої онтології», представленої на здобуття наукового ступеня кандидата технічних наук, використовуються у навчальному процесі кафедри «Інформаційні системи та мережі» Національного університету «Львівська політехніка». Матеріали дисертаційного дослідження використовуються під час написання студентами курсових робіт, кваліфікаційних бакалаврських та магістерських робіт, а також під час викладання дисциплін «Математична лінгвістика», «Організація баз даних та знань» та «Мови опису онтологій».

Зокрема, у навчальному процесі використовуються запропоновані О.В. Лозинською:

- загальні принципи синтаксичного аналізу речень (дисципліна «Математична лінгвістика» для студентів освітньо-кваліфікаційного рівня «бакалавр», що навчаються за напрямом 6.020303 «Філологія», тема 6 «Комбінаторна лінгвістика. Породжувальні граматики»);
- методи та засоби інформаційного моделювання предметної області, створення таблиць бази даних, проектування логічної структури реляційної бази даних (дисципліна «Організація баз даних та знань» для студентів освітньо-кваліфікаційного рівня «бакалавр», що навчаються за напрямом 6.050101 «Комп'ютерні науки», тема 2 «Проектування бази даних реляційного типу»);
- методи автоматизованого наповнення онтологій (дисципліна «Мови опису онтологій» для студентів освітньо-кваліфікаційного рівня «магістр», що навчаються за напрямом 8.05010104 «Системи штучного інтелекту», тема 1 «Основні визначення. Моделі онтологій, онтологічні системи. Системи і засоби подання онтологічних знань»);

Результати кандидатської дисертаційної роботи Лозинської О.В. подані та опубліковані у науковому виданні: *Лозинська О.В. Освітні інформаційні технології для глухих та слабкочуючих. Напрацювання та досвід Національного університету "Львівська політехніка" : препринт / О.В. Лозинська, М.В. Давидов, О.В. Вербич. – Львів: Видавництво Львівської політехніки, 2013. – 44 с.*

Директор ІКНІ,
 д.т.н., професор

М.О. Медиковський

Завідувач кафедри ІСМ,
 д.т.н., професор

В.В. Литвин

Професор кафедри ІСМ, д.т.н.

Р.М. Камінський



А К Т

**використання наукових результатів Лозинської Ольги Володимирівни,
отриманих у ході кандидатських дисертаційних досліджень за темою
«Система комп'ютерного перекладу української жестової мови з використанням
граматично доповненої онтології»
у науково-дослідній роботі
кафедри інформаційних систем та мереж**

Комісія у складі: голови комісії – начальника науково-дослідної частини, к.т.н., доцента Жук Л.В. та членів комісії – завідувача кафедри інформаційних систем та мереж, д.т.н., професора Литвина В.В., завідувача відділу науково-організаційного супроводу наукових досліджень, к.т.н. Лазько Г.В. та заступника начальника планово-фінансового відділу Чулой Т.М., цим актом підтверджують, що результати дисертаційної роботи Лозинської О.В. на тему «Система комп'ютерного перекладу української жестової мови з використанням граматично-доповненої онтології» використано у науково-дослідній роботі фінансованій Міністерством освіти і науки України «Математичне та програмно-алгоритмічне моделювання процесів трансляції жестів у текст для спеціалізованих комп'ютерних систем» (№ держреєстрації 0111U001222), що виконувалась на кафедрі інформаційних систем та мереж.

Отримані автором результати використано:

- при розробленні основних правил обходу дерев синтаксичного розбору речень української словесної мови та української жестової мови;
- при розробленні граматично-доповненої онтології української словесної та української жестової мов;
- при розробленні методу семантично-синтаксичного розбору речень української словесної мови та української жестової мови;
- при розробленні методу наповнення онтологій з використанням предметно-орієнтованих мов;
- при розробленні методів перекладу анотованої української жестової мови.

Голова комісії:

Начальник науково-дослідної частини
к.т.н., доцент

Л.В. Жук

Члени комісії:

Завідувач кафедри інформаційних
систем та мереж, д.т.н., професор

В.В. Литвин

Зав. відділу науково-організаційного
супроводу наукових досліджень, к.т.н.

Г.В. Лазько

Заст. начальника
планово-фінансового відділу

Т.М. Чулой