

ПРОЦЕСИ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПЕРЕКЛАДУ УКРАЇНСЬКОЇ ЖЕСТОВОЇ МОВИ НА ОСНОВІ ГРАМАТИЧНО ДОПОВНЕНОЇ ОНТОЛОГІЇ

© Лозинська О. В., 2015

Підсумовано доробок автора з розв’язання задачі перекладу української жестової мови. Відомі лінгвістичні моделі та методи перекладу порівняно з альтернативним підходом, який розробив автор для перекладу української жестової мови з використанням граматично доповненої онтології. Запропоновано метод наповнення онтологій з використанням предметно-орієнтованих мов, що дало змогу розробити граматично доповнену онтологію. Описано метод перекладу на основі граматично доповненої онтології.

Ключові слова: українська жести́ва мова, машинний переклад, синсет, онтологія, граматично доповнена онтологія.

The article summarizes the author's contribution for translation Ukrainian sign language. We describe the model of translation used for foreign sign languages. Among these statistical models of translation, the rule-based translation models and ontology-based models are described. These models are compared with the alternative approach for translation Ukrainian sign language using grammatically-augmented ontology. The method of filling ontologies using domain-specific languages, allowing us to develop a grammatically-augmented ontology is described.

Key words: Ukrainian sign language, machine translation, synset, ontology, grammatically-augmented ontology, domain specific language.

Вступ

Українська жести́ва мова (УЖМ) є природним способом спілкування людей з вадами слуху, тому необхідність отримання інформації жести́вою мовою для такої категорії громадян важлива. Сьогодні в Україні проживає близько 400 тисяч людей з вадами слуху. Для такої категорії громадян функціонує 59 спеціалізованих шкіл, 20 університетів (зокрема, Національний технічний університет України “Київський поліграфічний інститут”).

Жести́ва мова (ЖМ) – природна мова, основана на комбінації жестів, що використовуються для спілкування людей з вадами слуху. Жест виконується однією або обома руками в поєднанні з мімікою, а також в поєднанні з положенням корпусу тіла. Українська жести́ва мова є самостійною візуально-просторовою мовою та має свої правила побудови речень, відмінні від граматики української словесної мови (УСМ). Особливістю УЖМ є паралельне передавання інформації, яке можливе завдяки виконанню жестів обома руками, міміці та артикуляції. На відміну від цього в українській словесній мові передавання інформації відбувається лінійно (слово за словом). Поряд з УЖМ існує також калькована жести́ва мова (КЖМ), висловлювання якої повністю відповідають словесній мові [1].

У статті розглянуто інформаційні технології перекладу жести́вих мов у світі. Розроблено складові інформаційної технології перекладу української жести́вої мови на основі граматично доповненої онтології, яка не має аналогів у відомих наукових дослідженнях. Створені у процесі дослідження методи та програмні засоби опрацювання та перекладу української жести́вої мови можна використати під час розроблення програм перекладу з однієї мови на іншу, що полегшить спілкування осіб із вадами слуху між собою і з тими, хто не володіє жести́вою мовою.

Постановка проблеми

Для вирішення проблеми доступу до інформаційних ресурсів людей з вадами слуху розроблено словники та відеословники УЖМ, а також тренажери УЖМ. Проте немає ефективних засобів перекладу. Розроблення інформаційної технології перекладу української жестової мови є актуальним завданням. Така інформаційна технологія матиме велике соціальне значення, зокрема, дасть можливість особам з вадами слуху активно комунікувати із людьми, які не володіють жестовою мовою. Розроблення і комп'ютерна реалізація української жестової мови дадуть змогу широкому загалу людей вільно спілкуватися з людьми, які нечують.

Процес перекладу української жестової мови є складною проблемою, вирішення якої вимагає розв'язання низки задач, пов'язаних з усуненням неоднозначності слів, граматичним розбором жестових та словесних речень, збереженням та використанням знань експерта-перекладача. Інформаційна технологія перекладу української жестової мови має містити методи та засоби для розв'язання вищенаведених задач.

Відомі інформаційні технології перекладу жестових мов з використанням статистичних методів перекладу [2], методів на основі правил [3] та онтологій [4] не забезпечують належної якості перекладу, що зумовлює необхідність використання альтернативного підходу з граматично доповненою онтологією.

Для розроблення інформаційної технології перекладу української жестової мови реалізовано такі задачі: аналіз граматичних особливостей української жестової мови та сучасні підходи до створення систем перекладу жестових мов; декомпозицію задачі перекладу української жестової мови та проектування структури програмного комплексу; побудову системи правил для перекладу української жестової мови з використанням граматично доповненої онтології; розроблення методів інформаційної технології для перекладу української жестової мови із застосуванням граматично доповненої онтології; проведення експериментальних досліджень та виконання порівняльного аналізу отриманих результатів.

Автори визначили, що ключовими завданнями у процесі створення інформаційної технології перекладу української жестової мови є розроблення методів та засобів для ефективного перекладу УЖМ, оптимізація процесу наповнення граматично доповненої онтології, розроблення системи правил перетворення української словесної мови на жестову та навпаки.

Аналіз останніх досліджень та публікацій

Складність розроблення інформаційної технології перекладу української жестової мови посилюється відсутністю великих словників та корпусів УЖМ. Оскільки жестові мови використовують для подання інформації різні жести і вирази замість звуків, то для адаптування систем перекладу писемних мов для перекладу жестової мови необхідно розробити систему запису усіх елементів жестової мови, створити відповідні перекладацькі словники і системи граматичних правил для розбору жестової мови. Крім того, значна лінгвістична відмінність жестової мови від словесної ускладнює процес перекладу.

Огляд відомих інформаційних технологій перекладу жестових мов у світі показав, що найзастосовнішим для перекладу української жестової мови є машинний метод перекладу на основі правил та на основі онтологій.

Вчені Я. Бунджерот (J. Bungeroth) та Г. Ней (H. Ney) запропонували статистичний підхід до перекладу німецької жестової мови на текст та навпаки [5]. Процес перекладу складається з таких елементів: розпізнавання вхідних жестів та запису розпізнаних жестів у вигляді глосів, статистичного перекладу та запису вихідного тексту.

У роботі [6] розглянуто модель статистичного машинного перекладу з англійської мови на американську жестову мову. Складовими процесу 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. Схема системи статистичного машинного перекладу зображена на рис. 1.

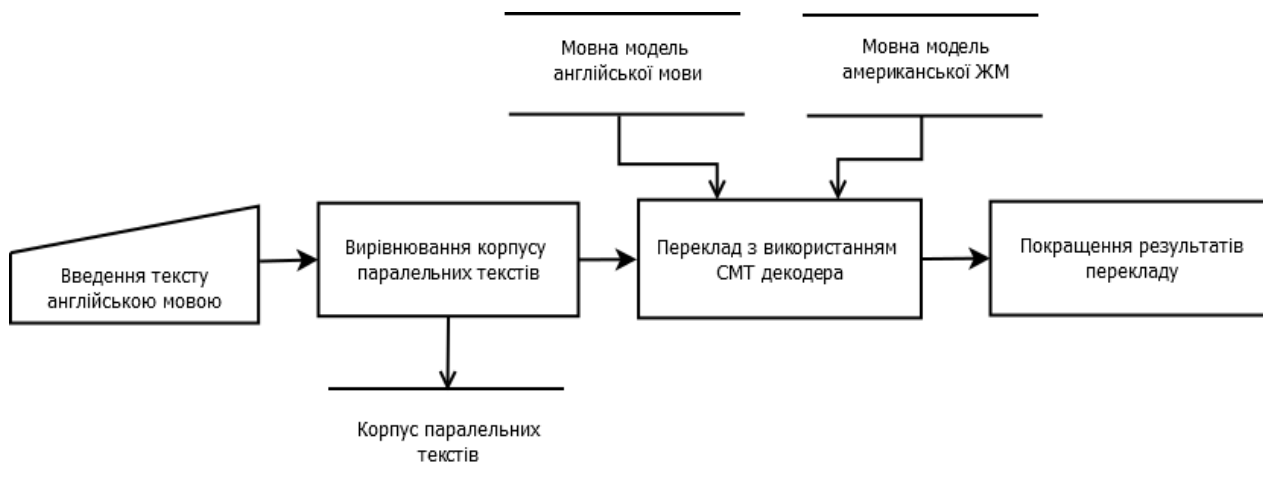


Рис. 1. Схема системи статистичного перекладу на американську жестову мову

Індійські вчені Т. Дасгупта та ін. [7] запропонували модель машинного перекладу на основі правил для трансформації тексту в індійську жестову мову. Процес перекладу на жестову мову складається з п'ятьох модулів: парсеру вхідного тексту, побудови LFG-структури, трансформації тексту з використанням правил граматики, генерування жестів, синтезу індійської жестової мови.

У роботі іспанських вчених Р. Сан-Сегундо та ін. [8] використано підхід на основі правил для перекладу розмовної мови на іспанську жестову мову. Переклад відбувається із залученням таких процесів: розпізнавання мовлення та його перетворення на послідовність слів, трансформації послідовності слів у послідовність жестів відповідно до правил граматики, анімації жестів. Процес перекладу здійснюється з використанням бази правил граматики іспанської жестової та словесної мов. Система перекладу на основі правил для іспанської жестової мови зображена на рис. 2.

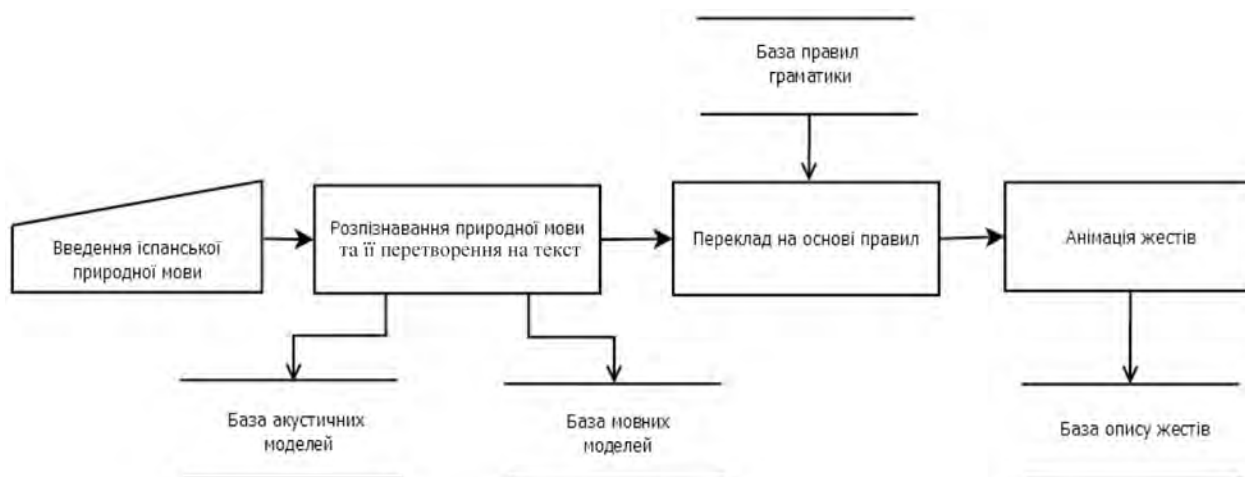


Рис. 2. Схема системи перекладу на основі правил для іспанської жестової мови

Інформаційну технологію перекладу польської жестової мови запропоновано у роботі [9]. Для польської жестової мови розроблено систему перекладу на основі правил, яка доповнена такими процесами, як синтаксичний та семантичний аналізи вхідного тексту, процес перетворення структури предикат-аргумент вхідного речення польською мовою на структуру предикат-аргумент вихідного речення жестовою польською мовою та процес анімації відповідних жестів. Семантичний аналіз речень здійснюється з використанням словника Semsyn та бази правил побудови структури предикат-аргумент. Схему системи перекладу для польської жестової мови наведено на рис. 3.

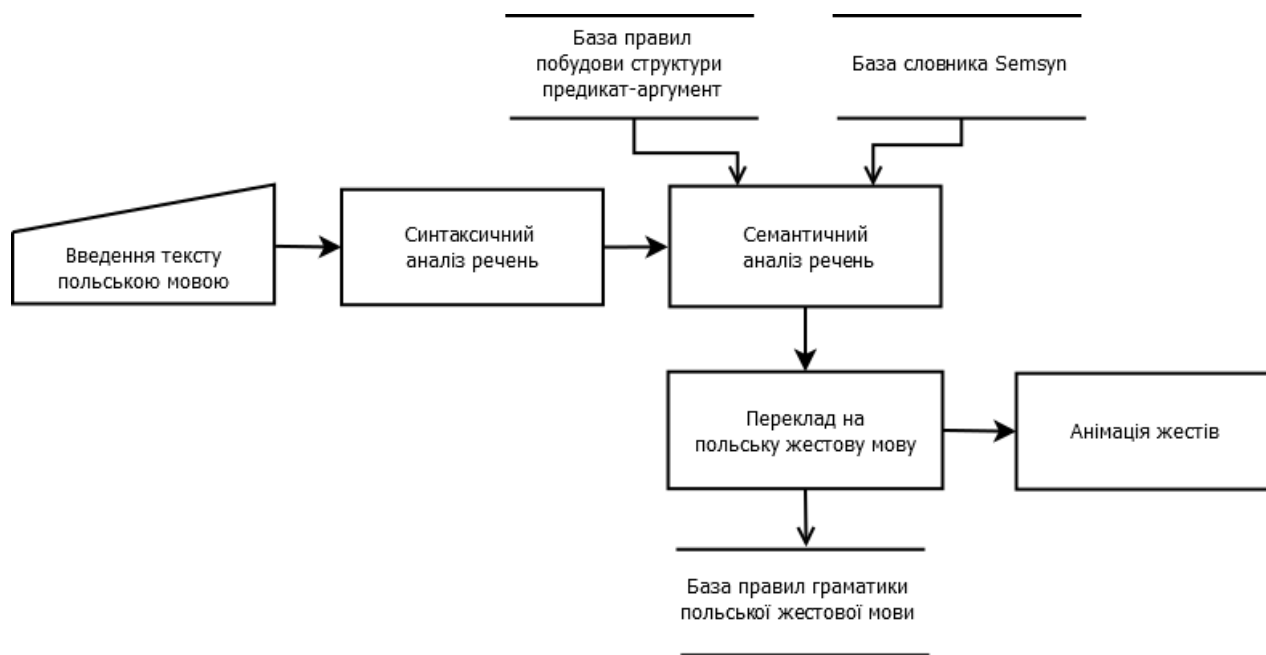


Рис. 3. Схема системи перекладу на основі правил для польської жестової мови

Статистичні системи машинного перекладу [5, 6] та системи перекладу на основі правил [7, 8] не використовують онтології предметної області.

Для перекладу арабського тексту на арабську жестову мову вчені А. Алмасуд та Х. О.-Халіф [4] запропонували модель перекладу на основі онтології. Процес перекладу складається з таких модулів: морфологічний аналізатор вхідного арабського тексту, трансформація тексту з використанням правил граматики, модуль перекладу на основі онтологій арабської словесної та арабської жестової мов (рис. 4). Трансформація речень відбувається із залученням бази правил граматики арабської жестової мови.

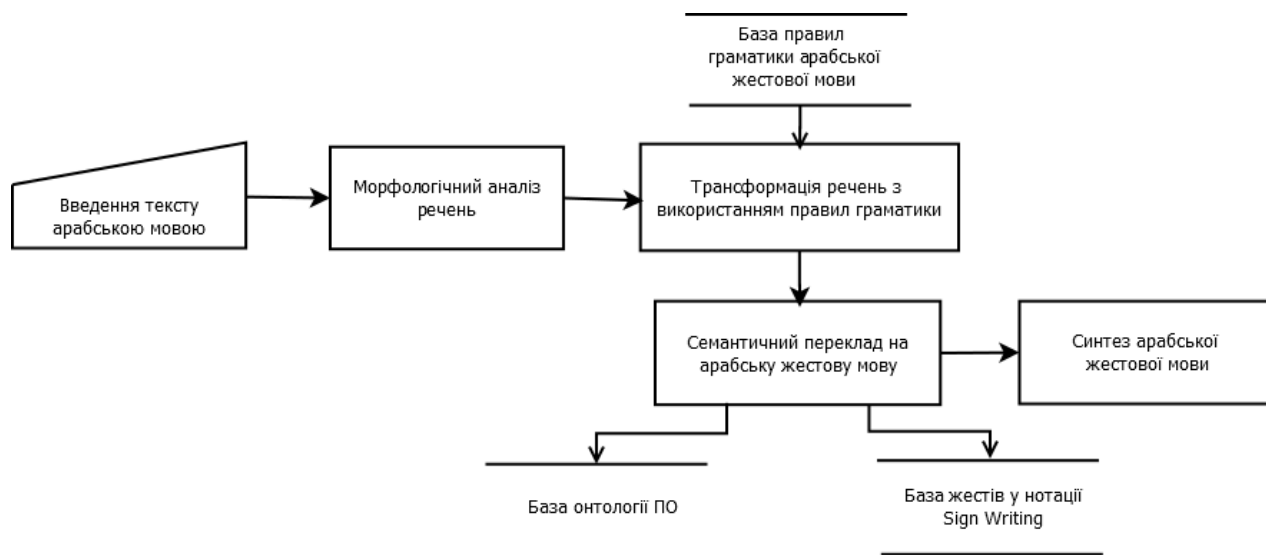


Рис. 4. Схема системи перекладу на арабську жестову мову з використанням онтології

Схему розробленої авторами системи перекладу з використанням граматично доповненої онтології зображено на рис. 5. На відміну від відомої системи перекладу на арабську жестову мову [4], яка також використовує онтології предметних областей, у системі перекладу з використанням ГДО спочатку здійснюється вибір параметризованих висловлювань з ГДО, а після цього

відбувається семантично-синтаксичний аналіз речення. Використання граматично доповненої онтології на першому етапі перекладу дає змогу надалі здійснити семантично-синтаксичний аналіз тексту та уникнути проблеми неоднозначного синтаксичного розбору речень. У системі перекладу (рис. 4) спочатку виконується морфологічний аналіз тексту, а потім вже використовується переклад із застосуванням онтології, що потребує аналізування речень для всіх можливих варіантів синтаксичного розбору з множини різних значень омографів.

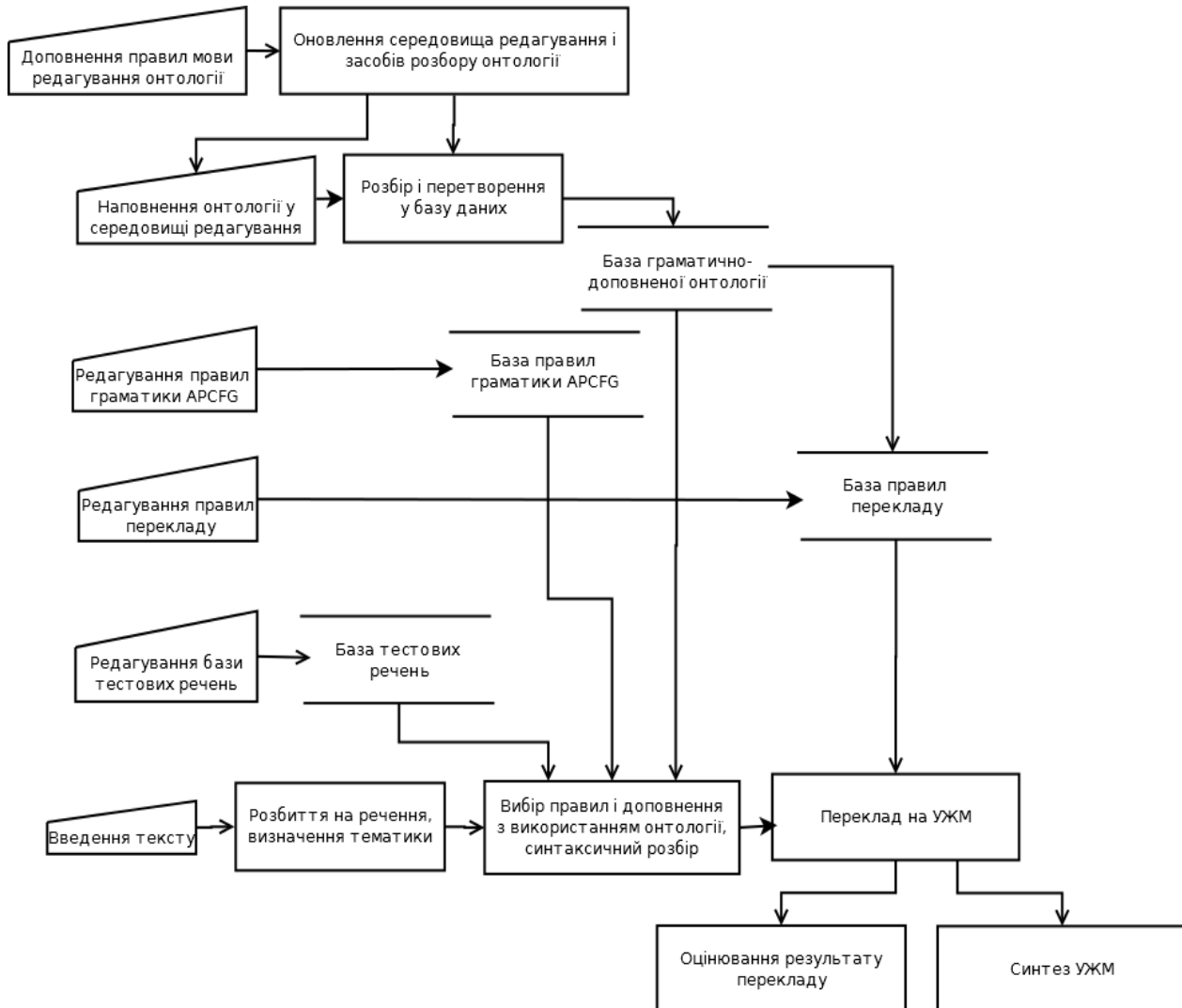


Рис. 5. Схема системи перекладу української жестової мови з використанням граматично доповненої онтології

Українські вчені Ю. В. Крак, О. В. Бармак та С. А. Романишин [10] запропонували інформаційну технологію автоматизованого перекладу з флективних мов на жестову мову. 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 вчені реалізували програмне забезпечення. За допомогою цієї технології отримано однозначний переклад без спотворення змісту в 100 % речень, узятих зі шкільної програми вивчення української мови жестів.

Ми дослідили інформаційні технології для перекладу жестових мов з використанням статистичних методів перекладу [2], [5], методів перекладу на основі правил [3] та методів перекладу на основі онтологій [4]. Проте ці інформаційні технології не підходять для перекладу УЖМ.

Процеси інформаційної технології перекладу української жестової мови

Інформаційна технологія перекладу жестової мови складається з таких процесів:

- наповнення граматично доповненої онтології [11];
- переклад із застосуванням правил [12, 13] та граматично доповненої онтології [14];
- тестування системи перекладу (якщо змінюються зв'язки в онтології, то це впливає на процес перекладу).

Процеси інформаційної технології зображено діаграмою використання на рис. 6.

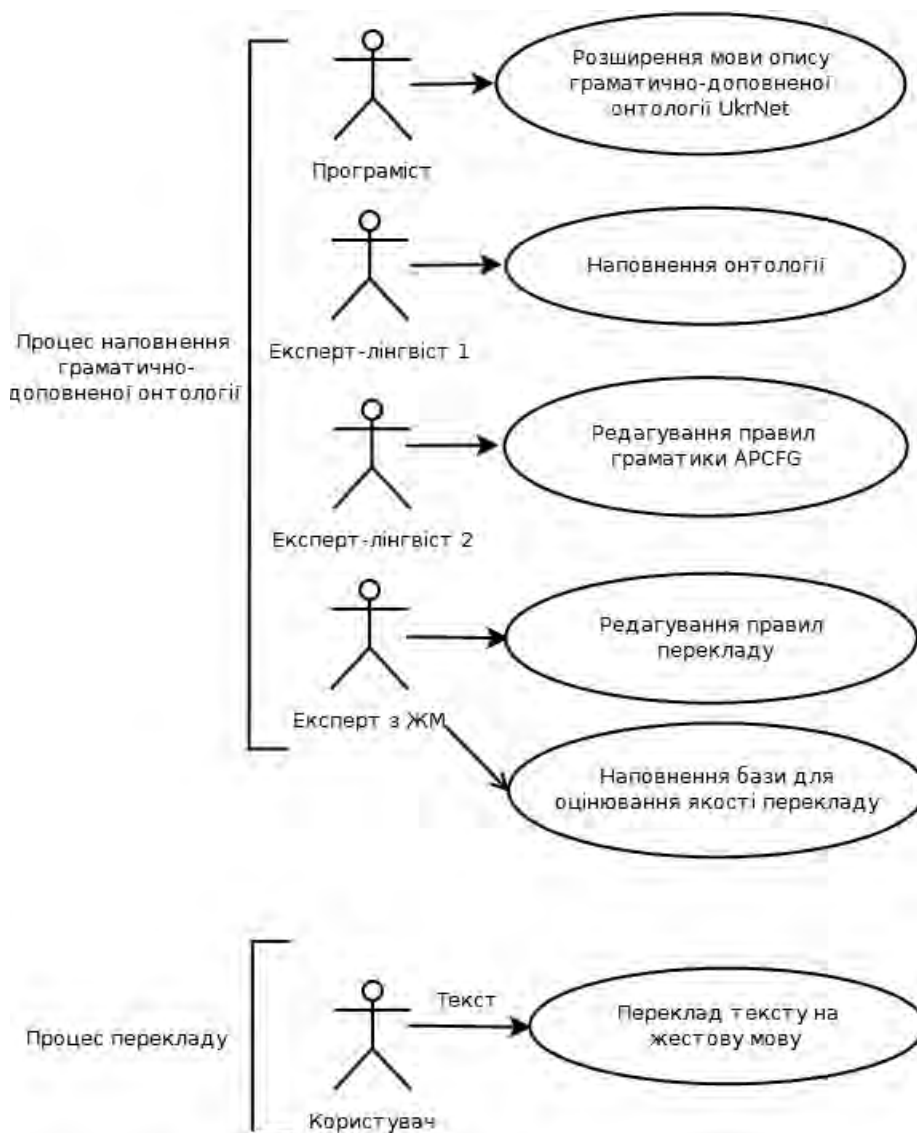


Рис. 6. Діаграма використання інформаційної технології перекладу

Основою для побудови мови опису граматично доповненої онтології стала математична модель граматично доповненої онтології, наведена у роботі [14]. Щоб уможливити введення в онтологію граматичних конструкцій української словесної та жестової мов, була розроблена предметно-орієнтована мова опису граматично доповненої онтології (Domain Specific Language, DSL). Термін “предметно-орієнтована мова” позначає мову програмування або моделювання, що застосовується для розв’язання конкретних задач у термінах, максимально наближених до певної предметної галузі [15].

Опис граматики мови UkrNet виконано із застосуванням додатка Xtext, що дало змогу використовувати середовище Eclipse для редагування онтології. Платформа Xtext на основі граматики DSL-мови забезпечує створення набору засобів для наповнення, редагування та

верифікації граматично доповненої онтології, що містить аналізатор коду мови, інструмент форматування коду, компілятор, редактор коду тощо. Використовуючи засоби Xtext, створено синтаксичний аналізатор і текстовий редактор для мови UkrNet [11].

Сьогодні граматично доповнена онтологія містить 1200 слів, пов'язаних такими галузями, як “Навчання”, “Природа”, “Подорож”, “Держава”, “Сім'я”, “Виробництво”, “Професії”, “Військо”, “Театр”, “Культура”, “Лікарня”. У ГДО наведено пояснення до кожного слова з використанням тлумачного словника української мови та словника жестів, занесено кожне слово у вигляді синсету (синонімічного ряду) та описано граматичні конструкції, в яких вжито це слово. Також у синсет введено граматичну конструкцію жестової мови.

Наприклад, розглянемо синсет “розповідати”. У словосполученні слово “розповідати” може означати “говорити”, “висловлювати думки”. Синсет для цього слова у граматично доповненій онтології:

```

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
жест говорити 1.ім_жест 2.ім_жест
жест розповідати 1.ім_жест 2.ім_жест
1: носій_знань
2: містить_знання
    
```

Детально з мовою опису ГДО можна ознайомитись у роботі [11].

На рис. 7 зображено блок-схему алгоритму перекладу із застосуванням граматично доповненої онтології.

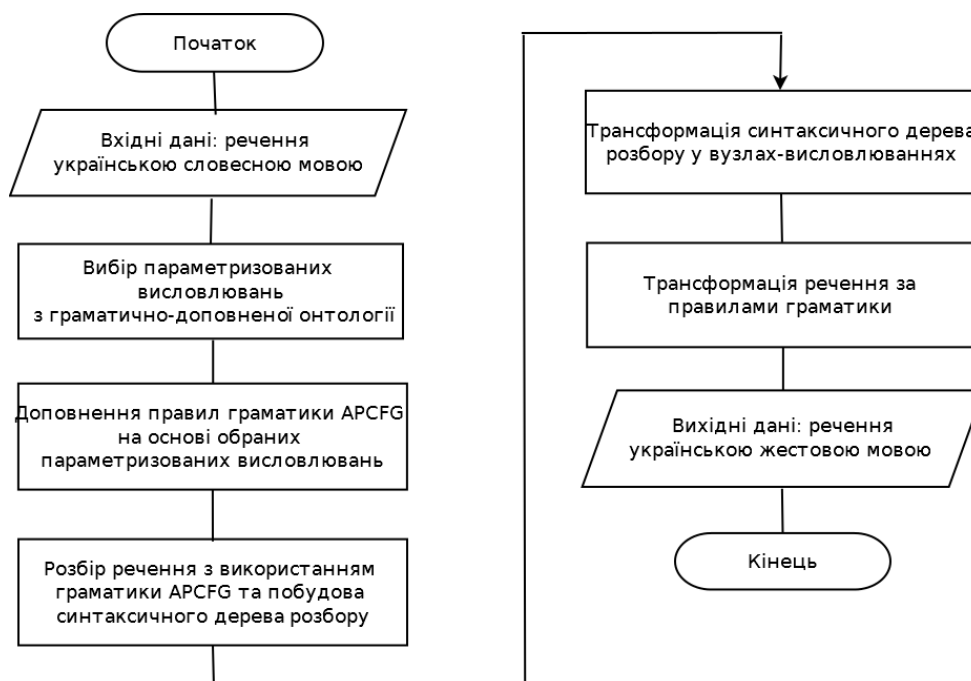


Рис. 7. Блок-схема алгоритму перекладу із застосуванням граматично доповненої онтології

На вхід подається речення українською словесною мовою. Далі програма здійснює пошук у граматично доповненій онтології та додає їх як правила виведення. Наступний етап полягає у розборі речення з використанням афіксної ймовірнісної контекстовільної граматики [11, 16]. У результаті розбору будується дерево семантично-синтаксичного розбору речення (рис. 8), коли розбір відповідає граматично доповненій онтології. Потім виконується трансформація семантично-синтаксичного дерева розбору речення за заданими правилами граматики УЖМ. На цьому етапі

встановлюється визначений відповідно до УЖМ порядок слів у реченні. Для перетворення речення УСМ на речення українською жестовою мовою розроблено спеціальну систему правил. Система складається з правил перекладу на жестову мову, правил порядку слів у реченнях відповідно до УЖМ [12, 13].

Дерево семантично-синтаксичного розбору для речення “Я розповідаю дітям казку”, коли розбір відповідає граматично доповненій онтології (вага більша за 1), зображено на рис. 8.

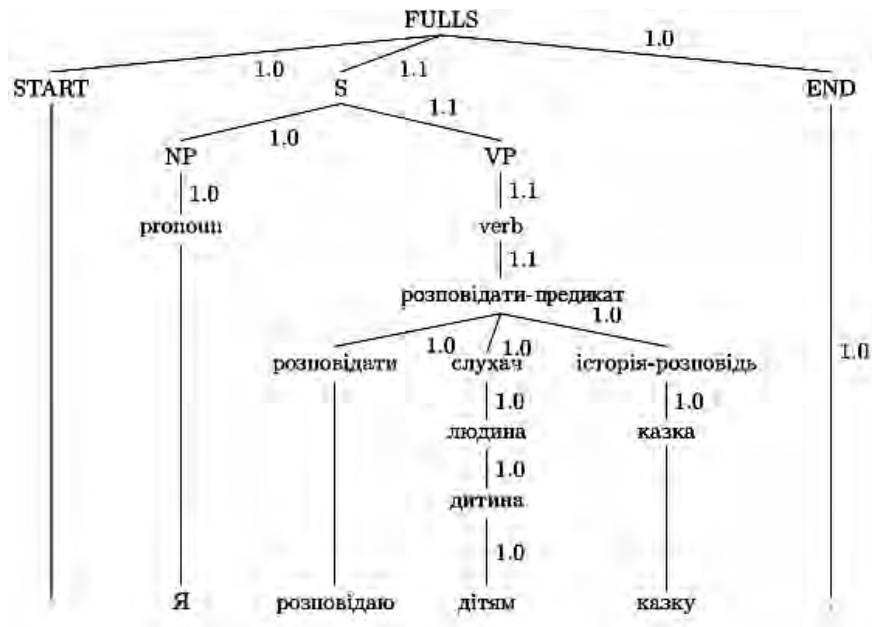


Рис. 8. Дерево семантично-синтаксичного розбору для речень “Я розповідаю дітям казку”

Оскільки у синсет входить граматична конструкція 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. 9. Можливість побудови дерев семантично-синтаксичного розбору для речень української словесної та жестової мов дає змогу здійснити однозначний переклад між цими мовами.

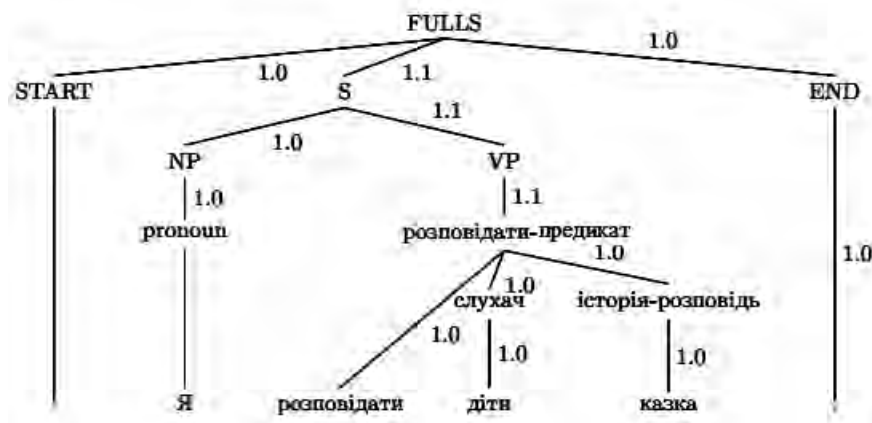


Рис. 9. Дерево синтаксичного розбору для речення “Я розповідати діти казка”

За допомогою відношень між поняттями у граматично доповненій онтології [14] на місця параметризованих змінних можна поставити лише правильні за змістом поняття. Відповідно до наведеної граматично доповненої онтології для речення “Я розповідаю столу казку” дерево синтаксичного розбору відповідає звичайному синтаксису речення (вага дорівнює 1) та зображено на рис. 10.

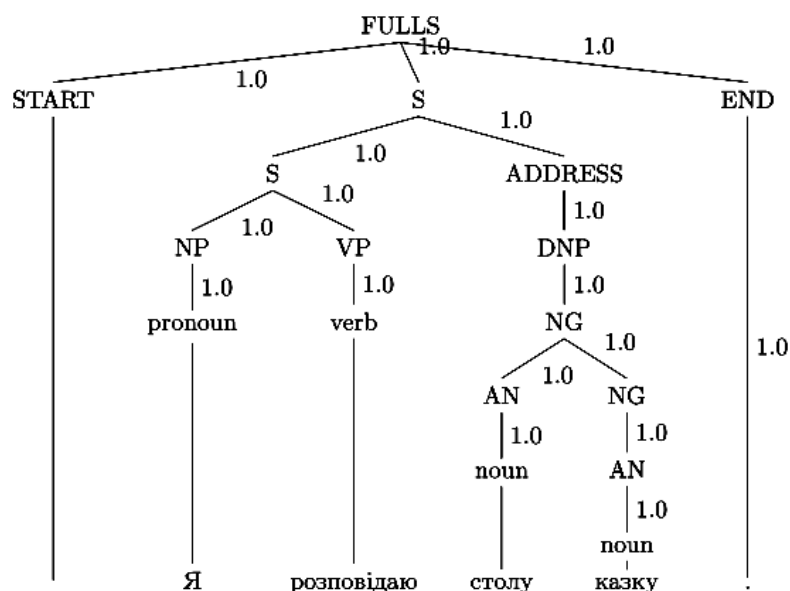


Рис. 10. Дерево синтаксичного розбору для речення “Я розповідаю столу казку”

У результаті проведених лінгвістичних досліджень української жестової мови описано правила обходу дерева семантично-синтаксичного розбору речень українською мовою для синтезу речень жестовою мовою. На основі прикладів дерев семантично-синтаксичного розбору речень української словесної мови визначено основні правила перекладу на українську жестову мову [12, 13].

Правила семантики, які створені з онтології для дерева семантично-синтаксичного розбору української словесної мови, зображеного на рис. 9:

<людина>[NCG] → <дитина>[NCG]
 <слухач>[NCG] → <людина>[NCG]
 <історія-розповідь>[NCG] → <казка>[NCG]
 <розповідати-синсет>[PNM] * 1.1 → <розповідати>[PNM] <історія-розповідь>[c4]
 <слухач>[c3]?
 verb[PNM] → <розповідати-синсет>[PNM]

Для української жестової мови правила семантики, які створені з онтології (рис. 11):

<слухач>[NCG] → <дити>[NCG]
 <історія-розповідь>[NCG] → <казка>[NCG]
 <розповідати-синсет>[PNM] * 1.1 → <розповідати>[PNM] <історія-розповідь>
 <слухач>?
 verb[PNM] → <розповідати-синсет>[PNM]

Розроблені методи для семантично-синтаксичного розбору речень, методи використання граматично доповненої онтології для перекладу української жестової мови в сукупності є основними елементами інформаційної технології перекладу УЖМ, описаної у статті.

Аналіз результатів

Для оцінювання якості системи перекладу української жестової мови виконано чотири експерименти, які полягали у перекладі тестувальних текстів:

- у першому експерименті для перекладу застосовувались статистичні методи перекладу та корпус паралельних текстів [17];
- у другому експерименті для перекладу використано лише правила перекладу [12, 13];
- у третьому експерименті переклад здійснювався з використанням словника концептів “УСМ-УЖМ” та правил перекладу [18];
- четвертий експеримент проводився із застосуванням правил та граматично доповненої онтології [15].

Результати оцінювання якості перекладу української жестової мови наведено у табл. 2 та на рис. 5.

Таблиця 2

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

Проведене оцінювання якості системи 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 української словесної мови на українську жестову мову до 93,2 %.

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

Висновки та перспективи подальших наукових розвідок

У статті досліджено актуальну наукову задачу комп’ютерного перекладу української жестової мови для людей з вадами слуху.

Для розв’язання цієї задачі розроблено методи та засоби інформаційної технології перекладу української жестової мови на основі граматично доповненої онтології. Виконано порівняльний аналіз відомих інформаційних технологій перекладу жестових мов у світі. Порівняно відомі лінгвістичні моделі та методи перекладу з альтернативним підходом, який розробила автор для перекладу української жестової мови з використанням граматично доповненої онтології.

Розроблено метод наповнення онтологій з використанням предметно-орієнтованих мов, що дало змогу розробити граматично доповнену онтологію. Описано метод перекладу на основі граматично доповненої онтології. Проведено експериментальні дослідження та виконано порівняльний аналіз отриманих результатів. Побудовано систему правил для перекладу української жестової мови із використанням граматично доповненої онтології.

Проведені дослідження показали високу ефективність застосування інформаційної технології перекладу української жестової мови на основі граматично доповненої онтології та можливість її використання в системах машинного перекладу.

1. Кульбіда С. В. Українська жестова мова як природна знакова система. *Жестова мова й сучасність: збірник наукових праць* / С. В. Кульбіда // К.: Центрпродрук. – 2009. – С. 218–239.
2. Stein D. *Hand in Hand: Automatic Sign Language to Speech Translation* / D. Stein, P. Drew, H. Ney, S. Morrissey, A. Way // *In Proceedings of Theoretical and Methodological Issues in Machine Translation (TMI-07)*. – Sweden. – 2007. – P. 214–220.
3. 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.
4. 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.
5. 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.
6. 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*. – 2011. – Vol. 8, Issue 5, № 3. – P. 65–73.
7. 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.

8. 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.

9. Suszczańska N. *Continuous Text Translation using Text Modeling in the Thetos System* / N. Suszczańska, P. Szmal, S. Kulików // *International Journal of Computational Intelligence.* – 2005. – Vol. 1, № 4. – P. 255–258.

10. Кривонос Ю. Г. Інформаційна технологія для моделювання української мови жестів / Ю. Г. Кривонос, О. В. Бармак, Ю. В. Крак, А. С. Тернов, Б. А. Троценко // *Штучний інтелект.* – 2009. – № 3. – С. 186–197.

11. Проект “UkrParser” [Електронний ресурс]. – Режим доступу: <https://github.com/mdavydov/UkrParser>.

12. Лозинська О. В. *Машинний переклад на основі правил для перекладу на українську жестову мову* / О. В. Лозинська, М. В. Давидов, В. В. Пасічник // *Інформаційні технології та комп'ютерна інженерія: міжнар. наук.-техн. журнал.* – Вінниця, 2014. – С. 11–17.

13. Лозинська О. В. *Застосування методу машинного перекладу на основі правил для перекладу з української жестової мови* / О. В. Лозинська // *Інформаційні системи та мережі. Вісник Національного університету “Львівська політехніка”.* – Львів, 2014. – № 783. – С. 368–375.

14. Лозинська О. В. *Математична модель граматично-доповненої онтології* / О. В. Лозинська, М. В. Давидов // *Вісник Національного технічного університету “ХПІ”.* – Х.: НТУ “ХПІ” – 2015. – № 11(1120). – С. 102–108.

14. Казакова А. С. *Методы и инструменты реализации предметно-ориентированных языков программирования* / А. С. Казакова. – М.: Вильямс, 2009. – С. 51–79.

16. Давидов М. В. *Грамматична корекція речень з використанням графу взаємовиключних гіпотез* / М. В. Давидов // *Вісник Національного університету “Львівська політехніка”.* – 2014. – № 783: *Інформаційні системи та мережі.* – С. 322–328.

17. 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. журнал.* – Харків: Харківський національний університет радіоелектроніки, 2013. – № 1(60). – С. 35–39.

18. Давидов М. В. *Інфологічна модель концептів української жестової мови* / М. В. Давидов, О. В. Лозинська, В. В. Пасічник // *Вісник Національного університету “Львівська політехніка” серія “Інформаційні системи та мережі”.* – 2014. – № 805. – С. 279–289.