

## ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

УДК 004.912

О.В. ЛОЗИНСЬКА, М.В. ДАВИДОВ, В.В. ПАСІЧНИК

Національний університет «Львівська політехніка», м. Львів

## МАШИННИЙ ПЕРЕКЛАД НА ОСНОВІ ПРАВИЛ ДЛЯ ПЕРЕКЛАДУ НА УКРАЇНСЬКУ ЖЕСТОВУ МОВУ

**Анотація.** Розглянуто сучасні досягнення у розв’язанні проблеми комп’ютерного перекладу з української словесної мови на українську жестову. Описано правила обходу дерева синтаксичного розбору речень українською мовою для синтезу речень жестовою мовою. На основі прикладів дерев синтаксичного розбору речень української словесної мови визначено основні правила перекладу на українську жестову. Розглянуто стверджувальні, заперечні й запитальні речення.

**Ключові слова:** українська жестова мова, корпус паралельних текстів, машинний переклад на основі правил, дерева синтаксичного розбору.

**Аннотация.** Рассмотрены современные достижения в решении проблемы компьютерного перевода с украинского словесного языка на украинский жестовый. Описаны правила обхода дерева синтаксического разбора предложений на украинском языке для синтеза предложений на жестовом языке. На основе примеров деревьев синтаксического разбора предложений украинского словесного языка определены основные правила перевода на украинский жестовый. Рассмотрены утвердительные, отрицательные и вопросительные предложения.

**Ключевые слова:** украинский жестовый язык, корпус параллельных текстов, машинный перевод на основе правил, дерева синтаксического разбора.

**Abstract.** Modern achievements in solving the problem of computer translation from Ukrainian Spoken Language into Ukrainian Sign Language are considered. Rules of traversing tree parsing sentences in Ukrainian for the synthesis of Sign Language sentences are described. On the basis of the examples of tree parsing sentences of Ukrainian Spoken Language, the rules for Sign Language translation are defined. Declarative, negative and interrogative sentences are considered.

**Keywords:** Ukrainian Sign Language, bilingual corpora, rule-based machine translation, parsing trees.

## Вступ

Розвиток сучасної науки, комп’ютеризація суспільства, використання мультимедійних та інтернет-технологій створили умови для розроблення комп’ютеризованих систем комунікації людей із особливими потребами, зокрема людей з вадами слуху. 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”). Актуальним завданням є розроблення програмних засобів, які допомагають у вивченні та перекладі української жестової мови.

Жестова мова (ЖМ) – це комунікативна система, висловлювання якої побудовані не на звуковій, а на жестикуляційно-мімічній основі (жести рук, артикуляція губ, міміка та емоції). Українська жестова мова (УЖМ) є природною візуально-просторовою мовою, яка використовується українським товариством глухих, заснованому в 1933 році. Ця мова незалежна від української словесної мови (УСМ), і не слід її плутати з калькованою жестовою мовою (КЖМ), яка повністю відповідає УСМ.

Дослідження української жестової мови розпочав Р. Г. Краєвський [1]. 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 більше 1500 жестів). На сьогоднішній день в Україні вивченням синтаксичних особливостей УЖМ займаються Н.Б. Адамюк, І.І. Чепчина [2].

## Постановка проблеми

Проблема перекладу українською жестовою мовою та навпаки є складною з двох причин. Перша з них — недосконалість жестових інтерфейсів комп’ютера, що ускладнює процес введення та виведення жестів. Друга причина — значна лінгвістична відмінність жестової мови від словесної. На відміну від калькованої жестової мови, яка повністю повторює порядок слів речення словесної мови, українська жестова мова має свій синтаксис та правила побудови речення.

Текстові та жестові словосполучення, пов’язані семантичними відношеннями, далеко не завжди перекладаються дослівно. Найчастіше зустрічається випадок, коли словосполучення, яке подає певний смисловий концепт, передається одним жестом або кількома жестами.

Метою статті є систематизація правил обходу дерева синтаксичного розбору речень українською мовою для синтезу речень жестовою мовою. На основі прикладів дерев синтаксичного розбору речень УСМ визначено основні правила перекладу на УЖМ. Розглянуто стверджувальні, заперечні й запитальні речення.

### Огляд відомих підходів

Серед відомих підходів до машинного перекладу жестової мови найбільш успішними виявилися підходи на основі методу трансферу та інтерлінгви [3], [4], [5].

Переклад на жестову мову передбачає синтаксичний розбір речень словесною мовою, усунення омографії, уточнення змісту та синтез правильних конструкцій жестовою мовою.

Не зважаючи на велику кількість наукових робіт, пов'язаних із синтаксичним розбором речень українською мовою [6], [7], [8], досі відсутні відкриті для використання програми автоматичного синтаксичного аналізу речень українською мовою.

Проблемами синтезу УЖМ займаються українські вчені Ю.В. Крак, О.В. Бармак та ін. з Інституту кібернетики ім. В.М. Глушкова НАН України [9]. При реалізації комп'ютерної технології для перекладу з української на жестову мову вчені стикнулися з проблемами, пов'язаними з наявністю двох жестових мов спілкування – калькуючої та української жестових мов, з тим, що немає однозначної відповідності слів жестам. У роботі [10] вчені запропонували механізм узагальнених граматичних конструкцій простих речень української та жестової мов на основі теорії лексикографічних систем [11] і побудували 764 узагальнених граматичних конструкцій із 6559 речень. У роботі не наведено результатів експериментів, які оцінюють правильність перекладу даної системи.

Не зважаючи на те, що лінгвістичні дослідження української жестової мови вже тривалий час проводяться науковцями, відсутні роботи, які описують усі аспекти граматики УЖМ та основних принципів перекладу на УЖМ.

Для англійської мови доступний цілий ряд синтаксичних аналізаторів [12], [13]. Граматичний розбір англійських речень може бути виконаний з допомогою граматики зв'язків (Links Grammar) [14], граматики синхронних приєднаних дерев (synchronous tree-adjointing grammar, STAG) [15], [16], лексико-функційної граматики (Lexical Functional Grammar, LFG) [5] тощо.

Для перекладу на жестову мову відомі різні методи, зокрема методи перекладу на основі прикладів [17], методи перекладу на основі правил [18], методи перекладу на основі даних [19] та статистичні методи перекладу [20], [21]. Іспанськими вченими Р. Сан-Сегундо та ін. [18] порівняно підходи на основі правил, на основі фраз та статистичний підхід для перекладу на іспанську жестову мову. За даними, наведеними у [18] найбільш точним виявився підхід на основі правил. Серед іноземних наукових розроблень для перекладу на жестову мову найбільш відомими є проекти ViSiCAST [22], TEAM [4], ASL Workbench [5], iCommunicator [23] та Artemis [24].

Для того, щоб розробити автоматизовану систему перекладу з української словесної мови на українську жестову та навпаки потрібно розробити синтаксичний аналізатор речень для обох мов, створити систему правил перекладу однієї мови на іншу, а також розробити корпус паралельних речень (УСМ – УЖМ) для порівняння якості перекладу.

### Правила синтезу речень українською жестовою мовою

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. Нижче розглянуто основні групи речень УСМ і наведено дерева синтаксичного розбору для кожної з них.

**Стверджувальні речення.** Дерево синтаксичного розбору для речення УСМ «Я дивлюсь цікавий фільм» зображено на рис. 1.

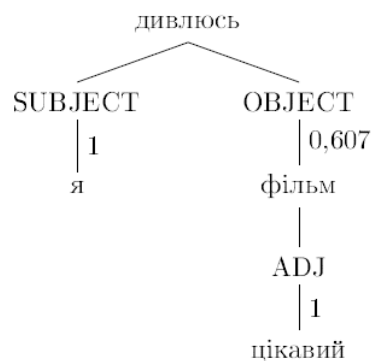


Рисунок 1 – Дерево синтаксичного розбору речення «Я дивлюсь цікавий фільм»

Оскільки українська словесна мова є флективною, то це ж саме речення ми можемо побудувати іншим чином: «Я цікавий фільм дивлюсь» або «Я цікавий дивлюсь фільм». Зміст речення при цьому не зміниться. Це не можна стверджувати щодо української жестової мови, в якій порядок слів строго визначений 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 таким чином: «Я дивитися фільм цікавий».

На основі правил граматики української жестової мови визначено правила перекладу, за допомогою яких отримуємо 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 для стверджувального речення УЖМ буде мати 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 предмету вживається інфінітив, а також допоміжні дієслова (Verb<sub>a</sub>, 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. 2). 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. Таким чином, уточнені правила для минулого та майбутнього часу будуть мати такий вигляд:

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

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

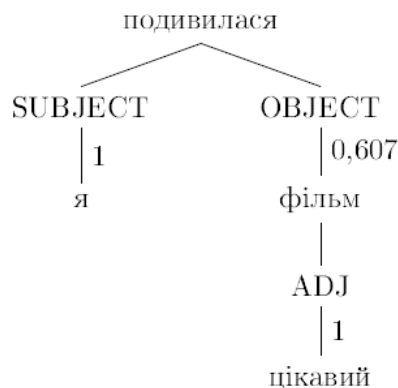


Рисунок 2 – Дерево синтаксичного розбору речення «Я подивилася цікавий фільм»

**Правило №3.** При побудові розповідного стверджувального речення прислівники використовуються після тих слів, до яких вони стосуються. Отже, набір правил для наведеного речення УЖМ буде мати такий вигляд:

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

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

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

Розглянемо дерево синтаксичного розбору для речення УСМ «Я нарешті прочитав цей текст» (рис. 3).

**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. 4).

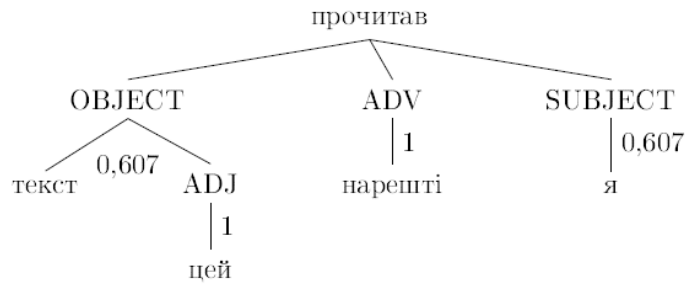


Рисунок 3 – Дерево синтаксичного розбору речення «Я нарешті прочитав цей текст»

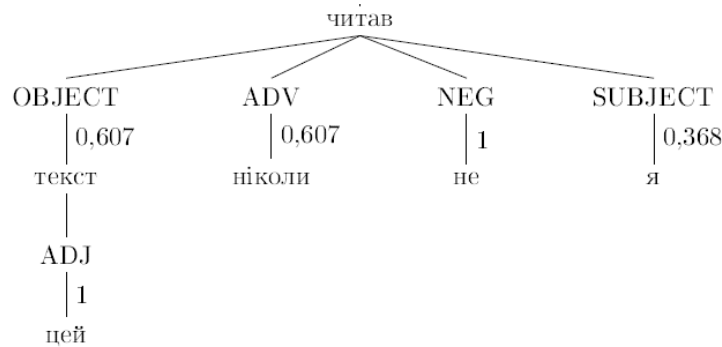


Рисунок 4 – Дерево синтаксичного розбору речення «Я ніколи не читав цей текст»

**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 для наведеного речення УЖМ буде мати 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»)[6]. 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. 5):

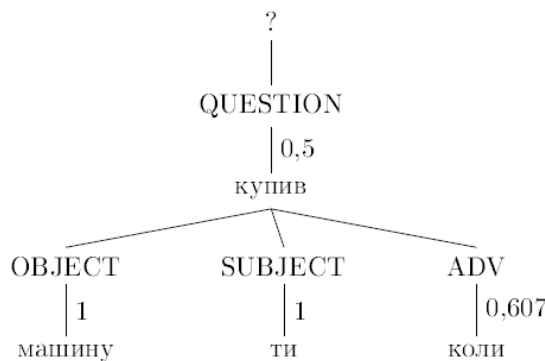


Рисунок 5 – Дерево синтаксичного розбору речення «Коли ти купив машину?»

**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 будемо використовувати. Отже, набір правил для питального речення УЖМ буде мати 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 і в українській словесній мові, та й 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. 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.

Розглянуто синтаксичний розбір окличного речення УСМ «Я нарешті прочитав цей текст!» (рис. 6):

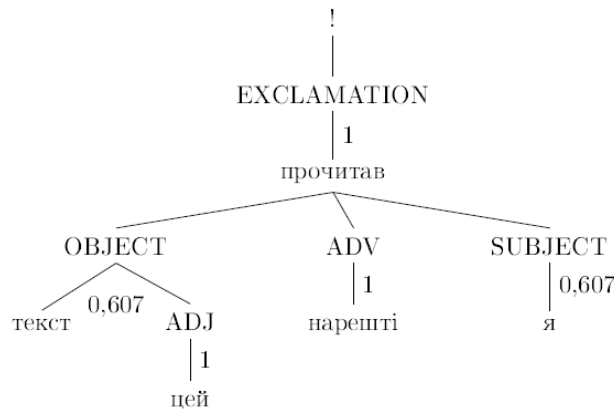


Рисунок 6 – Дерево синтаксичного розбору речення «Я нарешті прочитав цей текст!»

**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 будемо використовувати. Отже, набір правил для окличного речення УЖМ буде мати 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).**

#### Аналіз результатів

Для тестування перекладу було взято прості речення з корпусу паралельних текстів «Українська словесна мова – Українська жестова мова», який містить 220 речень, 932 слова УСМ і 839 слова УЖМ. Основні характеристики корпусу наведені у таблиці 1.

Таблиця 1 – Статистика корпусу паралельних текстів «Українська словесна мова – Українська жестова мова»

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

Переклад здійснювався на основі правил, виведених за допомогою дерев синтаксичного розбору речень. Для оцінки якості перекладу з української жестової на українську словесну мову використано

оцінки WER (word error rate) – оцінка помилок перекладу, PER (position-independent word error rate) – позиційно-незалежна оцінка помилок. Результати оцінки якості перекладу наведені у таблиці 2.

Таблиця 2 – Оцінка якості перекладу

	WER, %	PER, %
УСМ→УЖМ	42,83	18,35

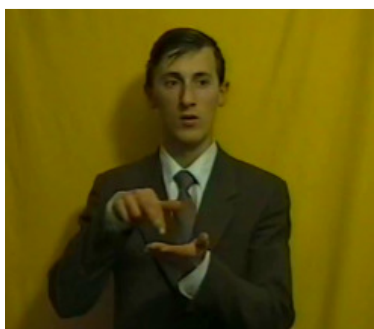
При перекладі з української словесної на українську жестову мову враховувалось, що усі іменники, займенники, прикметники та числівники в українській жестовій мові не мають закінчень, по яких можна ідентифікувати відмінок, рід та число, як у словесній мові. Також в УЖМ відсутні прийменники, а напрямок руху та розміщення жестів вже закладені у самих жестах. В українській словесній мові по закінченнях дієслів визначаються число, особа, рід, час, вид та спосіб, а у жестовій мові всі дієслова є у неозначеній формі дієслова.

Проте, враховуючи вищеописані правила, при перекладі виникали помилки. Нижче наведено кілька прикладів машинного перекладу з УСМ на УЖМ (табл. 3):

Таблиця 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. 7). 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

Рисунок 7 – Зображення жестів: а – «я йду»; б – «час йде»

### Висновки

Аналіз одержаних результатів показав доцільність використання машинного перекладу на основі правил для української жестової мови. Зокрема, були застосовані правила обходу дерев для синтезу речень українською жестовою мовою. Експерименти проводились на корпусі паралельних речень «Українська словесна мова – Українська жестова мова», який містить 220 речень. Подальшим дослідженням стане побудова правил перекладу з УСМ на УЖМ, враховуючи граматичні особливості української жестової мови. Результати цієї роботи можна використати під час розроблення програм перекладу з однієї мови на іншу, що полегшить спілкування осіб із вадами слуху між собою і з тими, хто не володіє жестовою мовою.

### Список літератури

1. Краєвський Р. Г. Мова жестів глухих / Р. Г. Краєвський. – К.: Рад. Школа, 1964. – 219 с.
2. Адамюк Н. Б. Синтаксичні особливості УЖМ: на прикладі простого речення / Н.Б. Адамюк, І.І. Чепчина // Жестова мова й сучасність: зб. наук. праць. – Випуск 4. – К.: Педагогічна думка, 2009. – С. 170-191.

3. Bangham J. A. Virtual Signing: Capture, Animation, Storage and Transmission – an Overview of the ViSiCAST Project / J. A. Bangham, S. J. Cox, R. Elliott, J. R. W. Glauert, I. Marshall, S. Rankov, M. Wells // *Speech and Language Processing for Disabled and Elderly People*, 2000.
4. 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. – pp. 293–300.
5. Speers d’A. L. Representation of American Sign Language for Machine Translation. PhD. dissertation, Graduate School of Arts and Sciences, Georgetown University, 2001.
6. Вихованець І. Р. Граматика української мови. Синтаксис / І. Р. Вихованець. – К., 1993. – 368 с.
7. Загнітко А.П. Основи українського теоретичного синтаксису. Частина 1 / А.П. Загнітко. – Горлівка, 2004. – 227с.
8. Дарчук Н.П. Автоматичний синтаксичний аналіз текстів корпусу української мови / Н.П. Дарчук // *Українське мовознавство*. – КНУ ім. Т. Шевченка, 2013.–№ 43.–С. 11-19.
9. 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» 3. – 2009. – С. 186-197.
10. Бармак О. Комп’ютерна технологія для перекладу з української на жестову мову / О.Бармак, С. Романишин // *Вісник ТНТУ*. – 2012. – Том 65. – № 1. – С.195-208.
11. Широков В.А. Інформаційна теорія лексикографічних систем [Текст] / В.А. Широков. – К.: Довіра, 1998. – 331с.
12. Програма CMU-парсер [Електронний ресурс]. – Режим доступу: <http://www.link.cs.cmu.edu/link/>
13. Програма LFG-парсер [Електронний ресурс]. – Режим доступу: <http://nlp.ioperm.org/lfg-parser.html>
14. Marshall I. Extraction of semantic representations from syntactic SMU link grammar linkages. In G. Angelova, editor, *Proceedings of Recent Advances in Natural Lanugage Processing*, Tzgov Chark, Bulgaria. – 2001. – pp. 154-159,
15. Shieber S.M. Synchronous tree-adjointing grammars / S.M. Shieber, Y.Schabes // *In Proceedings of the 13th International Conference on Computational Linguistics*, volume 3, Helsinki. – 1990. – pp. 253–258.
16. Zhao L. Machine Translation System from English to American Sign Language / L. Zhao, K. Kipper, W. Schuler, C. Vogler, N. Badler, and M. Palmer // *Association for Machine Translation in the Americas*. – 2000.
17. 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. – pp. 109–116.
18. 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. – pp. 2529-2532.
19. 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.
20. 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. – pp. 214-220.
21. Bungeroth J. Statistical Sign Language Translation / J. Bungeroth, H. Ney // *Proc. LREC 2004, Workshop proceedings: Representation and Processing of Sign Languages*. – Lisbon, Portugal. – 2004. – pp. 105–108.
22. Проект ViSiCAST [Електронний ресурс]. – Режим доступу: <http://www.visicast.co.uk/>
23. Проект «Communicator» [Електронний ресурс]. – Режим доступу: <http://www.icommunicator.com>.
24. Проект Artemis [Електронний ресурс]. – Режим доступу: <http://www.sim.int-evry.fr/Artemisю>

#### Відомості про авторів

**Лозинська Ольга Володимирівна** – аспірант кафедри інформаційних систем та мереж, Національний університет «Львівська політехніка».

**Давидов Максим Володимирович** – канд. техн. наук, доцент кафедри інформаційних систем та мереж, Національний університет «Львівська політехніка».

**Пасічник Володимир Володимирович** – д-р. техн. наук, Лауреат державної премії України в галузі науки і техніки, професор кафедри інформаційних систем та мереж, Національний університет «Львівська політехніка».